Privacy Discrimination: What It Is and Why It Matters

Luis Hillebrand
University of Geneva
luis.hillebrand@unige.ch

Sebastian Hermes
Technical University of Munich
sebastian.hermes@tum.de

Markus Böhm
Technical University of Munich
markus.boehm@tum.de

Abstract

We argue that online companies are able to exploit users' varying levels of privacy needs. We show that by employing data analytics methods on a comparatively small amount of data it is possible to predict how high information privacy concerns of specific users are. We argue that online companies might be able to introduce "privacy discrimination", in the sense that they might apply varying levels of privacy protection to users, based on their privacy concerns. Users indifferent about privacy could be presented with limited privacy options, adjusted terms and conditions or might be driven to disclose more personal information.

1. Introduction

One of the most vital streams of privacy research is the project of investigating peoples' privacy concerns [1, 2]. Researchers in information systems research have put huge amounts of work into conceptualizing privacy concerns as a way to operationalize previous, more diffuse concepts of when people care about privacy [1]. They have measured antecedents of privacy concerns on a personal level [3] and in regards to the relation of the individual to a corporation [4], and they have contextualized the phenomenon [5] and compared it across cultures [6]. Numerous theories draw on or extend privacy concerns [5, 7-11]. There are attempts to reintegrate all of this research into coherent frameworks [2, 12]. Put bluntly, the rationale for investing time into investigating privacy concerns can be summarized like this: If we understand peoples' privacy concerns, we can protect them from privacy threats where it is most needed. In particular the early literature on privacy concerns reflects this reasoning, justifying research on privacy concerns with privacy being "one of the most important ethical issues of the information age" [1, 13].

However, just like anything that is measured and managed, information about privacy concerns can be misused: Research on privacy concerns has revealed information about when, where, how, vis-à-vis whom etc. people are concerned about privacy. All of this information could potentially be used to extrapolate peoples' privacy concerns. This is especially troubling since there is a special emphasis in the research on the relation between privacy concerns and trust [14] and on individuals' willingness to share information with companies [15]. When it is possible to infer privacy concerns, it is also possible to control trust and information sharing, at least to some extent.

To capture this problem, we introduce the notion of privacy discrimination. It denotes the possibility that companies apply varying levels of privacy protection to different users. Most likely, this would happen on the basis of users' privacy concerns.

In order to explore the concept, we pose the following research questions: *What exactly is privacy discrimination? Is privacy discrimination technically feasible? Is investing in privacy discrimination be sensible for companies?*

In the remainder of this paper we develop a definition of privacy discrimination based on the notion of price discrimination in economics and drawing on attempts to define privacy. We show that it is technically feasible to infer users' privacy concerns from basic sociodemographic data about them. We discuss the implications of this technical possibility for individuals and online companies.

2. Theoretical background

2.1. Price discrimination

According to Stigler (1987), price discrimination occurs when the ratio of the prices of two similar products is different from the ratio of their marginal costs (Stigler, 1987). Stigler's seminal example is a book which is sold for $15 in the hardcover version but for $5 in paperback. This way, readers that are willing to pay $15 for a qualitatively superior book can do this, while readers that are only willing to pay less are not lost as customers. This example illustrates that prices can differ over time, but it is also possible for them to differ in space , by using early-bird discounts [16] or by identifying specific individuals directly [17]. While Stigler's definition is used widely, another prevailing definition is at least as relevant to this work. As Philips (1983, p. 5) puts it, "the usual answer: There is price discrimination when the same

URI: https://hdl.handle.net/10125/79943
978-0-9981331-5-7

commodity is sold at different prices to different customers" [18]. Although not entirely convincing even to Philips himself because of its inferior distinctive quality (different prices *and* different customers?), this definition is better suited for this work. Since privacy efforts do not have a direct equivalent to Stigler's marginal costs, it is more informative to apply Philips' definition here.[1]

There are three types of price discrimination Pigou [19]: First-degree or "perfect" price discrimination occurs when sellers charge a different price for each unit of a good adjusted to the maximum willingness to pay for this unit. Second-degree price discrimination or "nonlinear" pricing, occurs when sellers charge different prices depending on the number of units of the good bought, but not differing across consumers (e.g. quantity discounts). Third-degree price discrimination occurs when different purchasers are charged different prices, but each purchaser pays a constant amount for each unit of the good bought (e.g., student discounts).

The effects of price discrimination on a company, market and country level are topics of ongoing interest to economists. From a social welfare perspective, price discrimination does not have to be negative, but can have upsides [20], for example by fostering research and development (R&D) efforts [21]. However, regulation of prices can also delay launches of new medication or even deter the launch in specific countries [22].

Legally, the regulation of price discrimination differs across countries, but, as part of anti-trust regulations and other laws that secure healthy competition in markets, most legislations have at least some rules in place to regulate special forms of price discrimination or its application in specific areas . Many countries have rules against the "abuse of superior bargaining position", particularly with respect to long-term business relationships but also in cases where firms do not hold superior market power [23]. In general, these laws are designed to prevent firms from acquiring too much political and economic power, thus securing small firms' competitive possibilities.

Despite such regulations, in practice, price discrimination is not the exception but ubiquitous. It has been observed in various industries, most prominently with airlines [24] in Europe [25] and the US [26], but also for medicine [27] or academic journals [28]. Recently, online businesses have demonstrated their special ability to take advantage of price discrimination [17, 29, 30]. This seems to stem from their special expertise in Big Data, which has been shown to facilitate price discrimination [31].

2.2. Privacy and privacy concerns

The scientific discussion surrounding privacy is exceptionally vital. Regarding the number and variety of disciplines dealing with it, the diversity of perspectives taken and research questions raised and the methods used to tackle these questions, privacy is a truly amazing topic.

As a concept, defining privacy through concrete criteria has proven elusive and indeed undesirable [32], despite various attempts in the literature, see, e.g., [33, 34]. Based on this insight and on a more fuzzy understanding of privacy [35], Daniel Solove has proposed a taxonomy of privacy [36], organizing the activities where privacy issues can arise. Solove posits that privacy issues can arise as invasions in the form of intrusions and decision interferences towards individuals. Next, information collection in the form of surveillance or interrogations can harm individuals' privacy. Data holders can be responsible for privacy-sensitive information processing in the form of aggregation, identification, insecurity, secondary use, and exclusion. Finally, information dissemination in the form of confidentiality breaches, disclosure, exposure, increased accessibility, blackmailing, appropriation, and distortion can harm privacy. While the debate has certainly moved on since Solove published this taxonomy (2005) (cf. more recent reviews [13, 37]), these general pillars are still up to date. In information systems research, privacy concerns have become one of the major means to study privacy. Privacy concerns (often: Information Privacy Concerns, "IPC") are "concern(s) that individuals have with the information privacy practices of organizations, which could compromise the individuals' ability to control personal information" [1]. In the form currently used by most studies, IPC are able to measure individuals' concerns about the collection of information, errors regarding information, secondary use of information, and improper access to information [1].

Research on IPC is usually conducted against the background of the more general Antecedent-Privacy Concern-Outcome (APCO) Model [13]. This model suggests that there is a set of factors determining IPC (see below). IPC in turn can affect a set of outcomes, e.g., self-disclosure of information, specific privacy

[1] One might argue that in the case of privacy, marginal costs are equal for all users since it takes the same amount of technical cost to realize privacy levels. However, this is countered by the opportunity costs of high levels of privacy, which tend to reduce revenues.

behaviours [38], marketing effectiveness [39], the use of privacy measures [10] and settings [6] or the intention to use specific services [9]. Most studies follow this meta model either explicitly [referring to it by its name, e.g., 38] or implicitly [following the same modell without referring to directly, e.g., 39].

Numerous antecedent factors influencing IPC as antecedents have been studied. Regarding individuals, factors affecting privacy concerns include personality traits [3, 5], individual motives [8], prior experience with privacy breaches [9], perceived self-efficacy [10], perceived vulnerability [10], privacy consumption (i.e. whether someone reads privacy information) [40], personal roles [41], perceived justice regarding information privacy [41], perceived control [42] and individual IT culture [43]. Emotions also influence privacy concern: "Joy significantly enhances privacy protection belief and reduces privacy risk belief. Interestingly, fear was found to significantly influence privacy risk belief, but did not influence privacy protection belief" [44]. Moreover, perceived relevance of information requested and user awareness of the privacy policy incorporating fair information practice principles significantly increase privacy protection belief and reduce privacy risk belief [44].

Regarding organizations, factors affecting privacy concerns include trust in the company [39], boundary management and permeability (i.e., the possibility to decide on what to share with whom) [8], registration efficiency [4], perceived security level [10] and privacy settings [40]. In the relevant literature, particular attention has been paid to trust. There are different theories about the relationship of trust and IPC [14]; these describe trust as an antecedent of IPC [45], parallel to IPC as another construct [15], or as an outcome of IPC [3]. Also, trust can be conceptualized regarding the internet [7] or a specific company [46].

In general, privacy concerns are also affected by context [5] as well as culture [6]. More precisely, there seem to be positive effects of collectivism, uncertainty avoidance, and prior privacy experience on IPC [6]. In the context of location-based social networks, privacy control and privacy policies reduce privacy concerns [47]. Individuals' awareness of internet privacy legislation negatively influences privacy concerns, whereas previous privacy invasions do not [47]. Whether a location-based service is designed as push or pull has an impact on disclosure rates [48]. Regulation seems to lower perceived risks [48]. In e-commerce contexts, familiarity with a vendor helps to mitigate perceived privacy risks [49].

More specific theories making use of IPC are Privacy Calculus [7], Theory of Reasoned Action and Prospect Theory [5], Communication Privacy Management Theory [8], Social Contract Theory [9], Social Cognitive Theory, Protection Motivation Theory [10], and Social Exchange Theory [11].

2.3. Privacy discrimination

The preceding section reveals three rationales which motivate research on privacy concerns.

1. Privacy research investigates privacy concerns, because we want to protect people where they are concerned. In the APCO framework, this corresponds to investigating the relationships between Antecedents and Privacy Concerns. With better knowledge about antecedents, people can receive better protection.
2. Privacy research investigates privacy concerns, because privacy concerns are a predictor of outcomes in the APCO framework, such as peoples' willingness to share information, and we want to understand these outcomes.
3. As a vital sub-field of 2, we want to know about privacy concerns to be able to better infer peoples' privacy preferences. For example, according to the APCO framework, conversion rates and lower churn are important outcomes affected by individuals' privacy concerns.

While these accounts are justified, yet another view provides another perspective on the issue. In both of the strategies presented above, privacy concerns are only used as measurement. Information on privacy concerns is not perceived as constituting constructed facts that can change things in the world. Taking this view, it becomes clear that information on how privacy concerns work can be used in ways not intended by researchers investigating them. While there may be no intention to produce information that can be used for privacy discrimination in the way we describe it below in any of the strategies usually used to justify research on privacy concerns, their outputs still enable privacy discrimination.

Based on the definition of price discrimination in Philips [18] mentioned above, we define privacy discrimination as all organizational or individual practices that apply differing privacy levels to different users. Given this definition, privacy discrimination does not necessarily have to arise from privacy concerns, but depends heavily on technological implementations. One can imagine a doctor passing a patient's medical information to their partners based on an assessment of the privacy needs of the specific patient. Both in digital and in analogue form, privacy discrimination seems to be most likely when applied based on privacy concerns (cf. section

2.4, although using attributes such as age, gender, ethnicity etc. also are reasonable. Regarding Solove's (2005) taxonomy of potential privacy issues, privacy discrimination is likely to happen as an invasion (see *Example 2* in 2.4), as a decisional interference, during information collection (see *Example 3* in 2.4), as additional interrogation, or during information processing, e.g., when particularly intrusive computations are offered to users with low privacy concerns. Consequently, users with low privacy concerns are prone to privacy dissemination as a result of privacy discrimination.

Notice that privacy discrimination is *not* an instance of price discrimination but works in an analogous way. This analogy can be exploited to point out a few other similarities. Privacy discrimination seems to correspond to third-degree price discrimination, as different users have different privacy levels but the levels do not change over time. However, especially when customers have different concerns about different dimensions of privacy concerns (access, control etc. [1]), first-degree privacy discrimination becomes reasonable: Companies could change the privacy levels they apply from situation to situation. Furthermore, just like price discrimination, privacy discrimination should prima facie be considered value-neutral in order to enable a discussion about its advantages and disadvantages. Similarly, the economic implications of privacy discrimination are not obvious (see section 4, "Implications"). Privacy discrimination is able to exploit users' willingness to share information just as price discrimination is able to exploit customers' willingness to pay. Legally, regulation does not need to be specifically targeted at privacy discrimination but can still apply to it (just like anti-trust laws for price discrimination). For instance, several sections of the European General Data Protection Regulation (GDPR) might apply to privacy discrimination. Finally, because of their expertise in collecting and analyzing information, online companies are especially fit to apply privacy discrimination, in addition to price discrimination.

There is a similar concept to privacy discrimination in the literature: Personalized privacy [50] differs from privacy discrimination in the respect that personalized privacy lets users set their privacy preferences. In cases of privacy discrimination, privacy preferences (concerns) are inferred and privacy levels are applied based on these inferences, without explicitly asking the user. Hence, privacy discrimination has a paternalistic aspect that personalized privacy does not share. Given the importance of autonomy in modern political thought, this difference is fundamental.

2.4. Exploitation possibilities

Companies have various means to make use of privacy discrimination. All practices that enable them to increase profits based on more information about customers and more information sharing from customers can potentially be leveraged through privacy discrimination. Therefore, instead of trying to provide an exhaustive list, we provide three representative examples of how privacy discrimination could be exploited by firms. An overview is provided in Figure 1. To ensure relevance, the examples have to fulfil three criteria:

1. Adherence to the definition of privacy discrimination (cf. 2.3).
2. Technical feasibility, at least according to the results in the next chapter.
3. Financial plausibility with respect to existing business models (Section 4, "Implications of Privacy Discrimination").

	Users with high privacy concerns	Users with low privacy concerns
Ex. 1: Hiding settings/options	With whom would you like to share your photos? ○ Everyone ○ All friends ○ Selected friends	With whom would you like to share your photos? ○ Everyone ○ All friends
Ex. 2: Default settings	With whom would you like to share your photos? ○ Everyone ○ All friends ○ Selected friends	With whom would you like to share your photos? ● Everyone ○ All friends ○ Selected friends
Ex. 3: Asking for information	Please enter your age: _____ Please enter your gender: _____	Please enter your age: _____ Please enter your gender: _____ Please enter your sexual orientation: _____ Please enter your ZIP code: _____

Figure 1: Facsimiles of possibilities to exploit privacy discrimination

Example 1: Companies could hide certain options or whole sections of settings from users. In cases where users with low (predicted) privacy concerns were given fewer options to choose or no settings at all (for example, by not showing the possibility to select recipients of photos at all) compared to users with high privacy concerns, companies would be treating different users differently in regards to privacy. This represents an instance of privacy discrimination (fulfilling criterion 1.). Such discrimination is also technically implementable (criterion 2.): Based on the results from section 3., "Predicting Privacy Concerns", companies could direct users with predicted low privacy concerns between 1 and 2 to options like the ones depicted in the right column of Figure 1. Users with intermediate or high PCs (between 3 and 5) could be directed options like the ones depicted in the left column.

Example 2: Companies could preset privacy-related options for users. Consumers are especially susceptible to fall for biases [51] and nudges [52] when they are not highly concerned about a situation or want to move quickly, particularly in digital

contexts [53]. In cases where users with low (predicted) privacy concerns are given default choices with lower privacy levels than users with high privacy concerns, companies would be treating different users differently with respect to privacy (criterion 1.). This is also technically implementable (criterion 2.).

Example 3: Companies could ask users about more or less personal information at initial registration and potentially during use. In cases where users with low (predicted) privacy concerns were asked for more information (or more intrusive information, see, e.g., [54]) than users with high privacy concerns, companies would be treating different users differently with respect to privacy (criterion 1.). This is also technically implementable (criterion 2.).

3. Case study: Predicting IPC

In this section, we make the case for the technical feasibility of privacy discrimination. The purpose of this is to demonstrate that it is possible to predict peoples' privacy concerns with reasonable accuracy. In order to do this, we used experimental data to train a random forest classification model to predict participants' privacy concerns.

3.1. Materials & Methods

The dataset we used contained the answers of n=385 participants from a student sample collected for an experiment on privacy concerns. The dataset contained (amongst other variables that we do not use for this study) the age and gender of the participants as well as measures of their willingness to share information with a messenger service company and their privacy concerns about that company.

Participants' willingness to disclose information was collected using the items provided by Norberg, et al. [55]. We included the items that asked for the participants' willingness to disclose information on their personal pictures (disclosure1), cell phone number (2), location data (3), vacation time (4), address (5), and name (6). We chose these items because they are commonly shared to companies and websites such as social networks and are therefore relevant to our study. Online companies often have this kind of information and therefore can use it to infer users' privacy concerns. As there is some correlation between the behavioral intention to disclose information and actual information sharing (see section, 2.2.), these constructs serve as approximate measures for participants information sharing. Participants' IPC were collected using the constructs provided by Hong & Thong [12] [13] .

We used a random forest model to predict participants' privacy concerns. "A random forest is a classifier consisting of a collection of tree-structured classifiers […] where the [classifiers] are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input" [56]. We had 500 decision trees with 2 features tested per node. Varying the number of trees and variables at the nodes did not result in any significant differences in performance.

Random forests have several advantages over predictions based on regressions or other machine learning methods. Our main reasons for choosing a random forest for this application were the speed in learning [57] and classification [58], the possibility of determining the importance of variables used in classification [57], the ease of interpreting both results and the prediction process, and the fact that they are nonparametric and do not require specific scales or a unimodal distribution of the variables [57, 58]. Even more important for this specific case: In practice, decision trees are the second most important algorithm for data scientists (after regressions) and random forests the sixth most important, with the more important ones being either not applicable to our problem (cluster analysis, time series) or meta-algorithms, learning approaches etc. (ensemble methods). This ensures the practical relevance and applicability of our research.

3.2. Results

In reporting the results of the random forest in more detail, we follow the so-called A3 method. The A3 method as designed to report the results of various machine learning methods (*a*daptability) *a*ccurately and *a*ccessible, for practitioners as well as for researchers unfamiliar with machine learning [59]. Using the A3 package [60] we performed additional analysis[2] on the same data, in order to present more detailed information on the distribution of predictive qualities and the importance of the features in the models.

In Table 2, we provide the confusion matrix of the last round of cross validation. Table 3 suggests that most instances are classified correctly. Of those that are not, most are classified as a neighboring class. However, the model performs poorly for instances

[2] This led to minor differences between the results displayed in Table 2 and those displayed in the following tables. For example, in Table 4, we display decrease in Adj. R^2 for some features. These were not robust when exploring differing feature combinations (by exclusion of features) in cross-validation. These changes do not change any conclusions, qualitatively.

with IPC=1. This is probably due to the fact that there are only 29 (7.5%) such individuals in the data set. This is in line with the literature. Most of these instances are predicted as IPC=2. For an application of the method in practice this degree of accuracy is sufficient, and indeed favorable (see section 4, "Implications"): Instead of risking applying too low privacy protection to individuals that in fact are not concerned about privacy (i.e., have IPC=1), these individuals would have some basic privacy protection (according to IPC=2). This is precautionary.

Table 1: Confusion matrix

	1	2	3	4	5	Class error
1	2	18	2	1	1	0.916
2	9	48	16	7	4	0.429
3	1	25	28	3	6	0.555
4	1	8	9	51	28	0.474
5	0	1	1	24	42	0.382

Based on these results, we conclude that it is possible to predict peoples' privacy concerns solely by using their age, gender and willingness to disclose information[3]. Overfitting generally is not a problem in random forest models [56]. Still, in order to prevent random/statistical noise biasing otherwise-robust results, we used 5-fold cross-validation to assess our model. Table 3 reports the cross-validation. Column 2 displays the root mean square error (RMSE) of the prediction; column 3: mean error (ME); column 4: median error. Columns 5 and 6 display what percentage of the predicted instances are in a range of +/-.5, and 1, respectively, to users' true privacy concerns. Column 7 displays the out-of-bag error (OOB) of the model. Column 8 displays the R^2 for a random forest *regression* model, computed on the same data as the classification model we use throughout the study. This measure is only included as a service to readers not familiar with the OOB, which is better suited to evaluate classificatory models. As all error measures (RMSE, ME, Median Error) do not vary strongly between validation rounds, we conclude that our random forest predicts privacy concerns sufficiently well.

[3] Notice that this conclusion concerns only the possibility of the prediction of privacy concerns. We do not intend to make any statement about causality between personal information and privacy concerns, nor about which information is best to predict privacy concerns.

Table 2: Performance results of the random forest classification

Round of cross validation (k)	RMSE	ME	Median Error	In +/- .5 range	In +/- 1 range	OOB	Variance Explained
1	0.89	0.584	0	0.519	0.896	0.481	0.583
2	0.94	0.597	0	0.532	0.883	0.497	0.581
3	0.875	0.558	0	0.545	0.896	0.494	0.585
4	0.83	0.532	0	0.545	0.922	0.5	0.583
5	0.94	0.597	0	0.532	0.883	0.51	0.586
Mean	0.895	0.5736	0	0.5346	0.896	0.4964	0.5836

4. Implications of privacy discrimination

A larger sample size and more precise dataset would improve the confidence of our conclusion that privacy discrimination is feasible. In fact, we expect that a (considerably) higher prediction accuracy is achievable for online companies, given the following four factors:

1. Online platforms have better data science expertise than we have.
2. Online platforms have more data than we have.
3. Online platforms have higher computing capacities than we have.
4. Online platforms have the (kind of) data we have and more data, regarding online behavior, socio-demographic variables etc.

In conclusion, online businesses are able to predict people's privacy concerns with at least the accuracy and detail of this study.

4.1. Users' perspective

We draw on Osterwalder and Pigneur [61]'s characterization of customer relationships in e-businesses to discuss the implications of privacy discrimination on users. Based on their extensive research on business models, strategy, and processes, online companies need to balance the "feel and serve" with "trust and loyalty" in their customer relationship, and should use information strategies to determine the right balance.

Regarding the "feel and serve" dimension of customer relationship, privacy discrimination can be beneficial. Products that deploy privacy discrimination might present a more fluid customer experience. When users do not have to deliberately set options or think about with whom to share what information, they are spared time and stress. In

particular users with low levels of privacy concern might value the convenience of not being bothered with choices they do not care about. They might appreciate the possibility to be able to provide additional information that could increase product performance and improve their experience. Users with high privacy concerns might value feeling protected from the beginning of their relationship with a company without having to find settings to calibrate restrictively. On the other hand, when users try to find hidden settings or wonder why systems act the way they do (and the company did not communicate this as an instance of privacy discrimination) they might experience stress while trying to find and change settings. This would worsen the customer experience.

Regarding the "trust and loyalty" dimension, a similar picture emerges. Users might gain trust in a company when they see that settings are initialized according to their preferences potentially contributing to long-term customer loyalty. On the other hand, pre-set privacy settings that are not restrictive enough might cause customers to lose trust. Hence, from a user perspective, a cautious model for predicting privacy concerns seems favourable. This is reflected in the results, where users with minimal privacy concerns (=1) were classified as having privacy concerns =2.

Generally, skepticism about privacy discrimination is possible on three levels. First, users might disagree with specific decisions made based on privacy discrimination. When companies fail to pre-set options appropriately, users might be dissatisfied. Second, users might perceive privacy discrimination to be itself a privacy issue: Personal data are used to make sensitive inferences which might disseminate (Solove, 2005). Third, users might consider privacy discrimination to be a case of paternalism, which people tend to dislike, and be skeptical about it on these grounds. This analysis implies four possible views on privacy discrimination. Similar to privacy concerns, there might be *privacy discrimination concerns*. We discuss four idealized user types based on a matrix of these views (Figure 2).

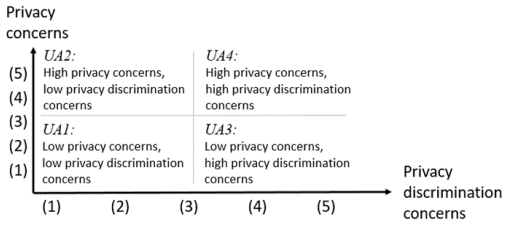

Figure 2: Matrix of possible attitudes towards privacy and towards privacy discrimination

User archetype (UA) 1: These users are not concerned about privacy and not concerned about privacy discrimination. They value the time savings offered by privacy discrimination.

User archetype 2: These users are concerned about privacy, but not concerned about privacy discrimination. They value the time savings of privacy discrimination, because they perceive it to protect their privacy concerns without interfering with their usage of the product.

User archetype 3: These users are not concerned about privacy but are concerned about privacy discrimination. They dislike privacy discrimination because they perceive it to be an invasion of privacy or they do not like having decisions made for them (paternalistically).

User archetype 4: These users are concerned about privacy and concerned about privacy discrimination. They are skeptical about sharing information in general, and they want to take care of their privacy themselves.

We hypothesize that *UA1* and *UA4* are especially common (i.e., privacy concerns and privacy discrimination concerns correlate). However, all four attitudes are plausible (as is a spectrum of intermediate attitudes).

4.2. Companies' perspective

Deploying privacy discrimination is most likely the decision of a company. To the best of our knowledge, there are no publicly known cases of actions based on privacy discrimination. We consider three possible explanations for this. First, companies might not yet have developed the means to engage in privacy discrimination. This seems unlikely, given the relative ease of determining our results (section 3). Second, cases of privacy discrimination might exist but not be publicly known. This is possible and would be concerning to advocates of privacy transparency. Third, companies might have deliberately chosen to not apply privacy discrimination. We center our discussion of companies' perspective on privacy discrimination on the question of why companies might so far have chosen not to deploy privacy discrimination.

While analyzing specific business processes is too fine-grained and analyzing business strategy too case-specific, analyses of the business models and stakeholders of online companies should reveal why we do not know of any cases of privacy discrimination. Different authors studying online business models propose different taxonomies and classifications. However, all agree on the importance of information for online businesses. It is especially important for marketing but also valuable in operations and other functions [62].

Abstracting from the peculiarities of specific business models proposed in these classifications, we work with a generalized, simplified model of online businesses (Fig. 3). We include the relevant stakeholders on the right because they are not strictly part of the business model (because they do not have value exchanges with the company) but they are important for the discussion. In this model, users give (or disclose) information to companies in order to receive information. Customers pay money for information. In some cases, users and customers are identical, while in others, like Facebook the user is not a customer. The user discloses personal information in order to see information about others and about companies, organization, and other groups. Customers can make use of the users' personal information to target advertising. For content providers or intermediaries (e.g. Netflix or eBay) the user is also a customer. Users give personal information and information about their needs and receive either information-as-a-product (content providers) or information-as-a-service, to facilitate physical exchanges (intermediary). In any case, information about customers is important to companies' success as is summarized by the notion of information as the "lifeblood of e-business". We indicate these relations with the grey block arrows in Fig. 3.

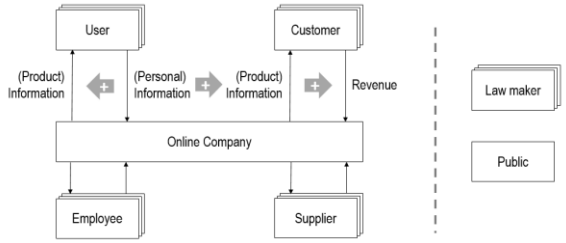

Figure 3: A generic model of online companies' business models, stakeholders and the importance of information[4].

By applying privacy discrimination, online companies are able to improve the way in which they capture the specific willingness of different customers to disclose information (analogous to the situation of price discrimination). Hence, online companies can collect more information (*Example 3*, section 2.4) and information marked by better quality (e.g., by replacing inferred ZIP codes with users' self-reported ZIP codes). The consequence of higher information quality or quantity can be better (product) information, for both users and customers. For example, in the case of Netflix, customers and users could receive recommendations that better fit their interests, increasing their satisfaction and activating better word-of-mouth effects. On the other hand, in the case of Facebook, users might receive news that fits their interests better and customers might receive more information, enabling them to improve the way in which they target their customers, in turn.

Examples 1 and *2* also show how product information might be improved for users. When users have less restrictive sharing settings, there is more potential information for other users to receive, potentially increasing the overall experience. This can lead to higher user numbers and less churn. Consequently, when (product) information for customers improves, revenues per customer increases and the number of customers also increases. All these measures increase the online company's revenue. Detrimental effects are possible as well. However, some cases where privacy discrimination is sensible from a financial perspective should exist. Turning towards other stakeholders of online companies explains the lack of publicly known cases of privacy discrimination. Privacy is an extremely sensible topic, and even more so in the public discussion. Facebook's Cambridge Analytica Scandal has shown that privacy invasions can seriously harm a company's image. Historical cases of popular outrage against companies [63] and research on corporate social responsibility [64] suggest that such cases can have serious financial consequences, both in compensation fees and in lost revenues because of image losses. These risks might keep managers from investing in privacy discrimination. Employees and suppliers perceive risks to the company as risks to themselves. Job losses or losses of important customers because of a suddenly worsened public image pose existential problems for both these groups. Employees could also see their own privacy at stake, when their company engages in practices that heavily exploit their users' privacy. Thus, while from a managerial point of view, privacy discrimination might offer potential increases in conversions and revenue, worries about employees and suppliers are likely to be in the way of applications of privacy discrimination. The authors of this work lack the expertise to assess the legal status of privacy discrimination. Online companies' managers might face the same situation. Both doubts about the legality and definitive knowledge about the illegality could explain missing evidence of privacy discrimination. In conclusion, we find that what has prevented the application of privacy discrimination so far are either legal hurdles or public concerns rather than a lack of technical feasibility.

[4] The grey block arrows indicate a causal impact, e.g., when users give more (or better) personal information to an online company, the product information that the company can give to customers is becoming more (or better).

5. Conclusion

All cases of practices where organizations or individuals apply differing privacy levels to different users are instances of privacy discrimination. This definition is open to varying concepts of privacy, but points to specific practices in collecting and processing information. We show that it is technically possible to infer peoples' privacy concerns with sufficient accuracy to base privacy discrimination on the results. While privacy concerns may not be the only basis for privacy discrimination, this is at least one case where privacy discrimination does work. There are minimal technical obstacles to implementing privacy discrimination.

Having argued that users can have varying attitudes towards privacy discrimination, we find that implementing privacy discrimination may be a sensible decision for some companies. This could lead to more and better data collection, which in turn could lead to better products, more customers, and higher revenues. However, public and legal hurdles may stand in the way of privacy discrimination, amplified by the concerns of employees and business partners.

6. References

[1] H. J. Smith, S. J. Milberg, and S. J. Burke, "Information privacy: measuring individuals' concerns about organizational practices," *MIS quarterly*, pp. 167-196, 1996.

[2] H. Yun, G. Lee, and D. J. Kim, "A chronological review of empirical research on personal information privacy concerns: An analysis of contexts and research constructs," *Information & Management*, vol. 56, no. 4, pp. 570-601, 2019.

[3] G. Bansal and D. Gefen, "The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online," *Decision support systems*, vol. 49, no. 2, pp. 138-150, 2010.

[4] P.-H. Hsieh and Y.-T. Hsiao, "Online User Intention to Select a Shared Account option on Multi-Service Platforms," in *PACIS*, 2016, p. 77.

[5] G. Bansal, F. M. Zahedi, and D. Gefen, "Do context and personality matter? Trust and privacy concerns in disclosing private information online," *Information & Management*, vol. 53, no. 1, pp. 1-21, 2016.

[6] B. Borena, S. Anteneh, F. Belanger, and D. Ejigu, "Conceptualizing Information Privacy Concern in Low-Income Countries: an Ethiopian Language Instrument for Social Network Sites Context," 2015.

[7] T. Dinev and P. Hart, "An extended privacy calculus model for e-commerce transactions," *Information Systems Research*, vol. 17, no. 1, pp. 61-80, 2006.

[8] A. Chennamaneni and A. Taneja, "Communication privacy management and self-disclosure on social media-a case of Facebook," 2015.

[9] H. Li, A. Gupta, J. Zhang, and R. Sarathy, "Examining the decision to use standalone personal health record systems as a trust-enabled fair social contract," *Decision Support Systems*, vol. 57, pp. 376-386, 2014.

[10] N. Mohamed and I. H. Ahmad, "Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia," *Computers in Human Behavior*, vol. 28, no. 6, pp. 2366-2375, 2012.

[11] D. Wilson, J. Proudfoot, and J. Valacich, "Saving face on Facebook: privacy concerns, social benefits, and impression management," 2014.

[12] W. Hong and J. Y. Thong, "Internet privacy concerns: An integrated conceptualization and four empirical studies," *Mis Quarterly*, pp. 275-298, 2013.

[13] H. J. Smith, T. Dinev, and H. Xu, "Information privacy research: an interdisciplinary review," *MIS quarterly*, vol. 35, no. 4, pp. 989-1016, 2011.

[14] F. Kehr, T. Kowatsch, D. Wentzel, and E. Fleisch, "Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus," *Information Systems Journal*, vol. 25, no. 6, pp. 607-635, 2015.

[15] C. L. Anderson and R. Agarwal, "The digitization of healthcare: boundary risks, emotion, and consumer willingness to disclose personal health information," *Information Systems Research*, vol. 22, no. 3, pp. 469-490, 2011.

[16] J. Dana, James D, "Advance-purchase discounts and price discrimination in competitive markets," *Journal of Political Economy*, vol. 106, no. 2, pp. 395-422, 1998.

[17] A. Acquisti, "Identity management, privacy, and price discrimination," *IEEE Security & Privacy*, vol. 6, no. 2, pp. 46-50, 2008.

[18] L. Philips, *The economics of price discrimination*. Cambridge: Cambridge University Press (in English), 1983.

[19] A. Pigou, "The Economics of Welfare. London: Macmillan and Co., ltd," 1920.

[20] M. A. Rappa, "The utility business model and the future of computing services," *IBM systems journal*, vol. 43, no. 1, pp. 32-42, 2004.

[21] S. Szymanski and T. M. Valletti, "Incentive effects of second prizes," *European Journal of Political Economy*, vol. 21, no. 2, pp. 467-481, 2005.

[22] I. M. Cockburn, J. O. Lanjouw, and M. Schankerman, "Patents and the global diffusion of new drugs," *American Economic Review*, vol. 106, no. 1, pp. 136-64, 2016.

[23] M. Wakui and T. K. Cheng, "Regulating abuse of superior bargaining position under the Japanese competition law: an anomaly or a necessity?," *Journal of Antitrust Enforcement*, vol. 3, no. 2, pp. 302-333, 2015.

[24] J. Stavins, "Price discrimination in the airline market: The effect of market concentration," *Review of Economics and Statistics*, vol. 83, no. 1, pp. 200-202, 2001.

[25] S. Giaume and S. Guillou, "Price discrimination and concentration in European airline markets," *Journal of Air Transport Management*, vol. 10, no. 5, pp. 305-310, 2004.

[26] S. Borenstein and N. L. Rose, "Competition and price dispersion in the US airline industry," *Journal of Political Economy*, vol. 102, no. 4, pp. 653-683, 1994.

[27] R. A. Kessel, "Price discrimination in medicine," *The Journal of Law and Economics*, vol. 1, pp. 20-53, 1958.
[28] E. A. Dyl, "A note on price discrimination by academic journals," *The Library Quarterly*, vol. 53, no. 2, pp. 161-168, 1983.
[29] A. Acquisti and H. R. Varian, "Conditioning prices on purchase history," *Marketing Science*, vol. 24, no. 3, pp. 367-381, 2005.
[30] A. Acquisti, "Ubiquitous computing, customer tracking, and price discrimination," in *Ubiquitous and pervasive commerce*: Springer, 2006, pp. 115-132.
[31] B. R. Shiller, *First degree price discrimination using big data*. Brandeis Univ., Department of Economics, 2013.
[32] C. Paine, U.-D. Reips, S. Stieger, A. Joinson, and T. Buchanan, "Internet users' perceptions of 'privacy concerns' and 'privacy actions'," *International Journal of Human-Computer Studies*, vol. 65, no. 6, pp. 526-536, 2007.
[33] F. Schoeman, "Privacy and intimate information," 1984.
[34] A. F. Westin, "Privacy and freedom Atheneum," *New York*, vol. 7, pp. 431-453, 1967.
[35] D. J. Solove, "Conceptualizing privacy," *Calif. L. Rev.*, vol. 90, p. 1087, 2002.
[36] D. J. Solove, "A taxonomy of privacy," *U. Pa. L. Rev.*, vol. 154, p. 477, 2005.
[37] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society*, vol. 3, no. 2, p. 2053951716679679, 2016.
[38] N. K. Lankton and J. F. Tripp, "A quantitative and qualitative study of facebook privacy using the antecedent-privacy concern-outcome macro model," 2013.
[39] A. Bleier and M. Eisenbeiss, "The importance of trust for personalized online advertising," *Journal of Retailing*, vol. 91, no. 3, pp. 390-409, 2015.
[40] F. Stutzman, R. Capra, and J. Thompson, "Factors mediating disclosure in social network sites," *Computers in Human Behavior*, vol. 27, no. 1, pp. 590-598, 2011.
[41] N. Zhang, C. Wang, and Y. Xu, "Privacy in online social networks," 2011.
[42] H. Xu and H.-H. Teo, "Alleviating consumers' privacy concerns in location-based services: a psychological control perspective," *ICIS 2004 proceedings*, p. 64, 2004.
[43] A. von Stetten, U. Wild, and W. Chrennikow, "Adopting Social Network Sites-The Role of Individual IT Culture and Privacy Concerns," in *AMCIS*, 2011.
[44] H. Li, R. Sarathy, and H. Xu, "The role of affect and cognition on online consumers' decision to disclose personal information to unfamiliar online vendors," *Decision Support Systems*, vol. 51, no. 3, pp. 434-445, 2011.
[45] R. Wakefield, "The influence of user affect in online information disclosure," *The Journal of Strategic Information Systems*, vol. 22, no. 2, pp. 157-174, 2013.
[46] H. Krasnova, N. F. Veltri, and O. Günther, "Self-disclosure and privacy calculus on social networking sites: The role of culture," *Business & Information Systems Engineering*, vol. 4, no. 3, pp. 127-135, 2012.
[47] L. Zhao, Y. Lu, and S. Gupta, "Disclosure intention of location-related information in location-based social network services," *International Journal of Electronic Commerce*, vol. 16, no. 4, pp. 53-90, 2012.
[48] H. Xu, H.-H. Teo, B. C. Tan, and R. Agarwal, "The role of push-pull technology in privacy calculus: the case of location-based services," *Journal of management information systems*, vol. 26, no. 3, pp. 135-174, 2009.
[49] C. Van Slyke, J. Shim, R. Johnson, and J. J. Jiang, "Concern for information privacy and online consumer purchasing," *Journal of the Association for Information Systems*, vol. 7, no. 6, p. 16, 2006.
[50] B. Liu *et al.*, "Follow my recommendations: A personalized privacy assistant for mobile app permissions," in *Twelfth Symposium on Usable Privacy and Security ({SOUPS} 2016)*, 2016, pp. 27-41.
[51] T. Gilovich, D. Griffin, and D. Kahneman, *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press, 2002.
[52] E. Selinger and K. Whyte, "Is there a right way to nudge? The practice and ethics of choice architecture," *Sociology Compass*, vol. 5, no. 10, pp. 923-935, 2011.
[53] M. Weinmann, C. Schneider, and J. vom Brocke, "Digital nudging," *Business & Information Systems Engineering*, vol. 58, no. 6, pp. 433-436, 2016.
[54] M. Morimoto and W. Macias, "A conceptual framework for unsolicited commercial e-mail: Perceived intrusiveness and privacy concerns," *Journal of Internet Commerce*, vol. 8, no. 3-4, pp. 137-160, 2009.
[55] P. A. Norberg, D. R. Horne, and D. A. Horne, "The privacy paradox: Personal information disclosure intentions versus behaviors," *Journal of consumer affairs*, vol. 41, no. 1, pp. 100-126, 2007.
[56] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001.
[57] D. Gao, Y.-X. Zhang, and Y.-H. Zhao, "Random forest algorithm for classification of multiwavelength data," *Research in Astronomy and Astrophysics*, vol. 9, no. 2, p. 220, 2009.
[58] N. Horning, "Random Forests: An algorithm for image classification and generation of continuous fields data sets," in *Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan*, 2010, vol. 911.
[59] S. Fortmann-Roe, "Consistent and clear reporting of results from diverse modeling techniques: the A3 method," *Journal of Statistical Software*, vol. 66, no. 7, pp. 1-23, 2015.
[60] S. Fortmann-Roe, "Package 'A3'," 2015.
[61] A. Osterwalder and Y. Pigneur, "An eBusiness model ontology for modeling eBusiness," *BLED 2002 proceedings*, p. 2, 2002.
[62] D. Chaffey, F. Ellis-Chadwick, R. Mayer, and K. Johnston, *Internet marketing: strategy, implementation and practice*. Pearson Education, 2009.
[63] D. Birsch and J. Fielder, "The Ford Pinto case: A study in applied ethics, business, and technology," 1994.
[64] M. S. Schwartz and D. Saiia, "Should Firms Go "Beyond Profits"? Milton Friedman versus Broad CSR 1," *Business and Society Review*, vol. 117, no. 1, pp. 1-31, 2012.