

Measuring Consumer Information Deficits in Transactions of Data-centric Services

Christian Flender, Martin Peters, Günter Müller

Institute of Computer Science and Social Studies

University of Freiburg

Friedrichstr. 50, 79098 Freiburg, Germany

{flender, peters, mueller}@iig.uni-freiburg.de

Abstract. *Recent success of data-centric services rests upon dual-valued transactions. In a first transaction services like online search and online social networks are offered for free (value 1). In exchange consumers provide data like clicking behavior or details about their identity. In a second transaction service providers turn into data providers selling or using data for targeting purposes (value 2), e.g. online advertising or credit inquiries. As consumers may not know about the way their data is used they are supposed to have information deficits. Once they find out about their data being used without consent they are likely to react with resentment and thus threaten market efficiency. We propose a model for measuring consumer information deficits in dual-valued transactions. Evidence of information deficits will support policy makers and software engineers in improving Transparency Enhancing Technology (TET) for effective signaling and screening.*

Keywords. Privacy, Data-centric Services, Information Deficits, Transparency

1 Introduction

Contrary to the assumption of complete market transparency consumers are far from perfectly informed about prices and quality of products and services [1]. Although in the digital economy transparency is significantly higher and transaction costs are significantly lower than in traditional markets, information deficits are prevalent. For instance, transactions on eBay are threatened by imposters, who take advantage of their anonymity and adopt strategies for manipulating reputation [2]. Information deficits pose a potential threat to trust among market participants and thus to market efficiency. Beyond classical eCommerce scenarios like online auctions recent internet services dealing with personal data equally rely on trusted relationships between service provider and service consumer. Data-centric services like online search and online social networks optimize eCommerce transactions. They support people in gathering information about products

and services less costly, e.g. offerings recommended from peer group members are likely to be more relevant. From this perspective data-centric services reduce transaction costs and information deficits.

However, in exchange for apparently free services consumers provide data like clicking behavior or details about their identity. Here information deficits are supposed to arise when service providers turn into data providers selling or using data for targeting purposes, e.g. online advertising or credit inquiries. As success of such services witnesses consumers generously give away data. Additionally, many studies observe a high appreciation for privacy¹ [4, 5, 6, 7, 8, 9, 10]. Privacy is the claim of individuals to determine for themselves how information about them is handled [11]. Information deficits about data usage prevent consumers from enforcing this claim. Accordingly, information is distributed asymmetrically among consumers and providers. On the level of individual behavior data protection laws counteract asymmetries. From a market perspective information deficits may result in economic imbalances necessitating regulation [12].

Do data-centric services cause inefficient market behaviour? To find out about states of informational imbalance we present an approach to measuring consumer information deficits in dual-valued transactions of data-centric services. We propose a design to survey consumers about their estimation of the price for data-centric services. A comparison of well-informed consumers with consumers being less informed about data usage is expected to reveal deviations between prices for services, i.e. differences in valuation of personal data. Consumers estimating the value of their privacy higher or lower compared to an equilibrium price are likely to shrink surplus and social welfare.

The paper is structured as follows. In the next section we motivate the problem of information deficits in markets of apparently free services. Having introduced the peculiarities of data as currency and privacy as absence of information deficits, in section 3, we come up with an approach to measuring information deficits.

¹cf. Flender and Müller (2012) for a thorough literature review on privacy decision making and the privacy paradox [3]

In section 4 we put forward potential avenues of designing Transparency Enhancing Technology (TET) as tools for effective signaling and screening. Elaborating on signaling and screening in the light of information deficits may bring to the surface new construction principles for TET. Finally, section 5 concludes the paper and looks into future research agendas.

2 Transactions of Data-centric Services

In free markets success of business models rests upon trusted relationships among market participants. Contrary to regulation and control, effective business transactions in free markets presuppose participants taking information deficits into account. Generally, trust among market participants does not require complete information, so why should there be a problem in markets of data-centric services?

There is empirical evidence for questioning trusted relationships among consumers and providers of data-centric services. Despite peoples' willingness to give away personal data in exchange for benefits a high appreciation of privacy is prevalent [4, 5, 6, 7, 8, 9, 10]. If consumers weren't concerned about their personal data, trust would be a matter of fact. Apparently, this is not the case. People do care about their privacy, in particular the way their data is used, sold, or handed over to third parties². A potential reason for privacy concerns is information about data usage being hardly available to consumers. Dual-valued transactions blur data usage (cf Figure 1). First, consumers use services for "free" and, knowingly or not, offer personal data in exchange (value 1). Second, service providers turn into data providers and make use of what consumers have told about themselves (value 2). At this point, privacy violation becomes a threat. Privacy is the claim of individuals to determine for themselves when, how, and to what extent information about them is communicated to others [11]. Consumers having information deficits about the way their data is used are supposed to have a lack of information about value 2. Once they find out about their data being used without consent, consumers are likely to react with resentment [13]. For measuring information deficits, we propose a model based upon demand and supply of data-centric services. We estimate information deficits in terms of consumers' expectations of the price for data-centric services from the background of their knowledge about data usage. From this estimation we put forward how consumer information deficits may diminish consumer surplus and social welfare.

Figure 2 shows an inefficient market equilibrium due to information deficits of consumers. Services supplied by providers are assumed to be perfectly elastic, i.e.

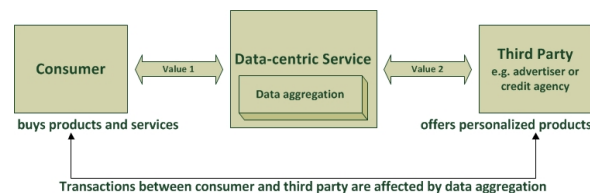


Figure 1: Dual-valued transactions of data-centric services. In a first transaction services like online search and online social networks are offered for free (value 1). In exchange consumers provide data like clicking behavior or details about their identity. In a second transaction service providers turn into data providers selling or using data for targeting purposes (value 2), e.g. online advertising or credit inquiries. As consumers may not know about value 2 they are supposed to have information deficits.

providers are willing to offer any quantity of their service for a fixed price p^* . This is a reasonable assumption since economies of scale are typical for markets of digital services where fixed costs are high and marginal costs tend toward zero [14]. The market price p^* is the price all users have to pay for using data-centric services. The prices p' and p'' reflect consumers' estimation of the price for services, i.e. the value of personal data consumers give in exchange for using services. If users were able to know the market price p^* , the resulting equilibrium would be (x^*, p^*) . This equilibrium is efficient. It maximizes social welfare on the market according to area $A + B + C$. However, as observed

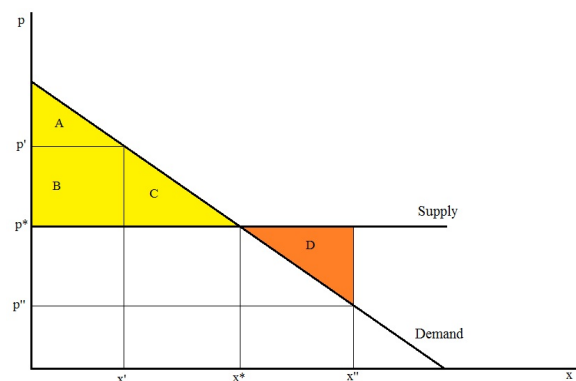


Figure 2: Information deficits about data usage prevent consumers from estimating the market price. Two scenarios are possible. First, consumers estimate the price less than the actual market price $p'' < p^*$. This may diminish consumer surplus according to area D . Second, consumers estimate the price higher than the actual market price $p' > p^*$. This may diminish consumer surplus according to area C . In both cases information deficits about data usage have an impact on market efficiency.

in several studies, consumers are concerned about their privacy despite generous data disclosure in exchange

²cf. <https://www.cdt.org/privacy/guide/surveyinfo.php>

for apparently "free" services. Privacy concerns nourish the assumption that consumers have information deficits about the actual price they pay for using services. Therefore, it is reasonable to assume that p^* is not known. Rather, we expect deviating prices p' and p'' . Two scenarios are possible:

1. Consumers estimate the price p' where $p' > p^*$. In this case, the market equilibrium is (x', p^*) . Consumers demand x' services for price p^* . Here the equilibrium is inefficient and consumer surplus may diminish according to area C.
2. Consumers estimate the price p'' where $p'' < p^*$. The market equilibrium is (x'', p^*) . Consumers whose valuation of privacy is less than p^* but more than p'' participate in the market due to their information deficits. Their surplus may shrink according to area D.

Figure 2 illustrates our approach to observing information deficits indirectly. Measuring the market price p^* of data-centric services directly from a supply-side perspective will be hard if not impossible for several reasons.

First, surveying providers about revenues earned from using customer data may reveal the market price determined by value 2. However, for providers what counts first and foremost is value 1, i.e. offering (free) services for better search results, improved socialization, or eased communication (cf. Figure 1). This is what personal data is used for in the first place. Keeping the way data generates value 2 secret appears rational from the point of view of competition. Providers are likely reluctant to inform customers, competitors or anyone else about the actual value of consumer data.

Second, estimating the market price based on spending efforts for data collection and analysis techniques may give a clue. The amount of investments in data collection mechanisms such as web mining and other Business Intelligence (BI) technologies may correlate with value 2. However, business analytics is a broad field spanning from material management, logistics, production and other operational business processes to strategic decision making. Finding relationships between value 2 and BI investments is a rather vague endeavour.

From difficulties of measuring prices directly we follow that it will be more promising to look at consumers' estimation of the price for data-centric services rather than estimating a market price from a supply-side point of view. If consumers actually have information deficits, then less informed users will estimate the price differently compared to better informed consumers. Thus it is not necessary to know the actual market price for personal data.

In the next section, we come up with a survey design to measure consumer information deficits. We divide participants into control group and treatment group. Both groups will be asked to estimate the

price for using data-centric services. The treatment group will receive additional information about how their data is used by providers. We want to find out if consumers' valuation of personal data differs between the two groups. From this we are able draw conclusions regarding information deficits and potential consequences such as diminishing consumer surplus and social welfare.

3 Measuring Consumer Information Deficits

There are several studies investigating transactions in markets where two groups of participants are brought together via intermediaries. In such two-sided markets (or two-sided networks) examples for platforms realizing dual-valued transactions are credit cards connecting clients and merchants, search engines connecting searchers and advertisers, or operating systems connecting users and developers [15, 16]. For one side of the market products and services are free or subsidized, from the other side the platform generates revenues. Similar to network externalities on markets for information technology [14], cross-side network effects span across dual-valued transactions of data-centric services. They can be positive, e.g. the more searchers spend their attention, the more advertiser are attracted. Or they are negative. For instance, the more advertisements are shown on search result pages, the more searchers prefer fewer advertisements. Beyond cross-side network effects, free or subsidized services haven't been illuminated yet in the context of consumer information deficits about data usage.

According to Bellotti and Sellen (1993), we refer to data usage as a concept embracing four candidates of consumer information deficits [17]: (1) capture, (2) construction, (3) accessibility, and (4) purpose.

1. **Capture** describes information deficits about the kind of information gathered by data-centric services. Examples are clicking behavior, personal identity, purchasing history, messages sent, profiles visited, etc.
2. **Construction** describes information deficits about the way data is processed. For instance, many merchants build customer profiles and infer potentials for cross- and up-selling from similar shopping behaviors [18].
3. **Accessibility** describes information deficits about social groups, e.g. the public, governmental agencies, advertisers, or application services like web tracking tools having access to personal data.
4. **Purpose** describes information deficits about the intent behind data usage. Knowing what others intend to do with personal data, e.g. commercial

trade, personalized offerings, etc., may influence the value of privacy.

The four criteria for potential information deficits were developed for resolving transparency issues in the context of ubiquitous computing and computer-supported cooperative work. Data-centric services are intrinsically cooperative; they connect participants in realizing platforms on two-sided markets. For instance, search providers and social networks collect and make use of consumer data by connecting advertisers, searchers, and avatars respectively. For both business models consumers may have deficits regarding the kind of information collected (capture), the way their data is enriched with other sources (construction), the target group interested in their data (accessibility), as well as the intent behind their data being used, traded, or handed over (purpose). The proposed categories are general enough to instantiate several concrete scenarios, e.g. social networks or online search. They also set the scope for further studies and provide a terminology to render them comparable.

In the following we outline structure and content of two basic questions for measuring information deficits. After that we describe how we prepare our survey in terms of experimental conditions. Here we apply our framework for preparing the treatment group. Our questionnaire embraces two questions. Each of them addresses consumers' estimation of the price for data-centric services in a slightly different manner.

1. Your service provider collects your personal data according to its data protection policy. How much a month would you charge a third party for making use of that data?
2. Another provider offers the same service as your current provider but does not allow third parties to make use of data. How much a month would you be willing to pay for the service?

Questions address consumers' estimation of the price for data-centric services either by asking for the value of personal data positively, i.e. data usage (Q1), or negatively, i.e. data protection (Q2). In order to provide a set of choices for making privacy judgments each participant will receive multiple options for assessing personal data, for instance, a range of values between €10 and €100. For measuring consumer information deficits the real monetary value is not the crucial point. Rather what counts is a comparable measure applicable to both treatment and control group. We are interested in finding a significant deviation between the two groups. To one (treatment) group information about data usage is provided. Another (control) group does not receive any information. Information provided to the treatment group is meant to emulate the absence of information deficits. Accordingly, the treatment group is better informed and our hypothesis assumes that information makes a difference to how consumers will

judge the value of their personal data. There are two options for preparing the survey.

1. Providing information prior to handing out the questionnaire. Unlike the control group, the treatment group receives a short text which describes the kind of data collected (e.g. clicking behaviour), the way it is processed (e.g. aggregation with other sources), who has an interest (e.g. advertisers), and, what is the objective (e.g. targeting).
2. Providing information in line with the questionnaire. Treatment group and control group both receive a questionnaire with the same type of questions. However, for each question information about capture, construction, accessibility, and purpose is provided. The reason for binding information to each question is that we want to find out about each single variable, their interdependencies, and impact on privacy judgments. For instance, consumers may be more sensitive to the target group who is interested in their personal life than the particular type of data collected.

Eventually, if we get a significant deviation between both groups, we have evidence of an inefficient equilibrium as shown in Figure 2. In the following we consider mechanisms to reduce information deficits.

4 Signaling and Screening

In the previous section we proposed an instrument for measuring consumer information deficits. Consumers who estimate the price for data-centric services higher or lower compared to the actual market price are supposed to have information deficits about the way their data is used. Information deficits pose a threat to market efficiency as consumers are likely to react with resentment once they find out about their data being used without consent. Undesired consequences may be diminishing consumer surplus and social welfare (cf. Figure 2). As theoretical considerations suggest, avoiding such consequences necessitates the reduction of information deficits.

In this section we discuss signaling and screening as potential means for reducing information deficits. For signaling the better informed market participant - the provider, legislator, or trusted third party - sends a signal to reduce deficits (heteronomy). Consumers actively screening for information about data usage reduce deficits in a self-guided fashion (autonomy). Transparency Enhancing Technology (TET) embraces signalling and screening mechanisms. In contrast to Privacy Enhancing Technology (PET), TET appears more apt for protecting privacy in transactions of data-centric services. PET is usually designed to support data avoidance [19]. The objective is to keep the amount of information required for trusted business

transactions as low as possible. Moreover, PET is typically designed to support 1:1 and 1:n transactions in which consumers and providers have full control over their devices and data, e.g. E-mail communication and client-server interaction. Most data-centric services, however, require n:m interactions where data dissipation lies at the heart of their business model. Users of social networks usually increase benefits the more and the often they disclose information about themselves and their daily life. Attracting attention sustainably requires the opposite of data avoidance, i.e. generous data disclosure. Therefore, in the context of data-centric services privacy appears to be less a matter of data avoidance and access control than transparency and control of data usage.

Several effects of TET as a mean for signaling and screening have been put forth. Böhme (2009) studies the impact of TET on social behavior, in particular effects on diversity and conformity [19]. By means of a multi-period game with heterogeneous preferences two conflicting hypotheses stand against each other. Both hypotheses acknowledge TET as an effective instrument for reducing information deficits. The first, however, leads to "mainstream" behavior where participants conform to shared information about data usage. For instance, if people were to know about moving in a "risky" area has an impact on their credit scoring, they would collectively avoid passing through that area and thus conform to a shared social norm. The second hypothesis claims that transparency supports diversity because without transparency people are herded together by uncertainty and fear. As an overall result of the simulation study transparency tends to foster conformity although effects are rather weak.

Hildebrandt (2007) proposes TET as a mean to counter threats of profiling [20]. As a form of data usage, profiling, i.e. the construction and application of consumer profiles, is prevalent in transactions of data-centric services. The author argues that protection of personal data alone isn't sufficient anymore and must be extended towards protection against the undesired application of profiles and the creation of transparency rights regarding group profiles. PET brings into focus the hiding of data, i.e. anonymization, and the use of pseudonyms. Increasingly, data is tracked and applied for profiling but citizens will need more than the possibility of opting out. With TET consumers can render accessible and assessable the profiles that may affect their life. In consequence TET may prevent consumers residing in a golden cage where providers anticipate preferences before consumers become aware of them.

In their early work on the design of privacy, Belotti and Sellen (1993) acknowledge that interactions between participants mediated via technology are prone to breakdowns. Breakdowns occur due to inadequate feedback signaling the way data is used. Also an inability to control access to data [17], a lack of screening, compromises trust. The authors consider privacy as a

central design issue in its own right and come up with a framework for TET that we adapted for data-centric services.

5 Conclusion and Future Work

From difficulties of data avoidance in transactions of data-centric services we conclude two crucial points to be determined in order to inform policy makers and software engineers. First, there must be evidence of information deficits and inefficient markets respectively. Second, required evidence is given, transparency mechanisms for signaling and screening may counteract information deficits. Likewise, effectiveness of TET is in need of studies delivering evidence. For both (1) information deficits and (2) TET effectiveness, we came up with a survey design and generalized model. The model embraces principles for legitimizing and designing TET in times where data avoidance isn't really an option anymore.

We are about to conduct a first survey with consumers of social networks. Required our hypothesis turns out to be true, studies in other domains will follow. Also we plan to investigate available mechanisms for signaling and screening from the background of their promise to reduce consumer information deficits in transactions of data-centric services.

References

- [1] Akerlof, G.: The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics* (1970) 488–500
- [2] Reitzenstein, I., Peters, R.: Assessing robustness of reputation systems regarding interdependent manipulations. *E-Commerce and Web Technologies* (2009) 288–299
- [3] Flender, C., Müller, G.: Type Indeterminacy in Privacy Decisions: The Privacy Paradox Revisited. In: *Proceedings of the 6th International Symposium on Quantum Interaction (QI 2012)*, Paris School of Economics, ENS, France, Springer (2012)
- [4] Norberg, P., Horne, D., Horne, D.: The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs* **41** (2007) 100–126
- [5] Awad, N., Krishnan, M.: The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS quarterly* (2006) 13–28
- [6] Acquisti, A., Grossklags, J.: Privacy and rationality in individual decision making. *Security & Privacy, IEEE* **3** (2005) 26–33

- [7] Berendt, B., Günther, O., Spiekermann, S.: Privacy in e-commerce: Stated preferences vs. actual behavior. *Communications of the ACM* **48** (2005) 101–106
- [8] Sayre, S., Horne, D.: Trading secrets for savings: How concerned are consumers about club cards as a privacy threat? *Advances in Consumer Research* **27** (2000) 151–155
- [9] Sheehan, K., Hoy, M.: Flaming, complaining, abstaining: How online users respond to privacy concerns. *Journal of Advertising* (1999) 37–51
- [10] Spiekermann, S., Grossklags, J., Berendt, B.: E-privacy in 2nd generation e-commerce: privacy preferences versus actual behavior. In: *Proceedings of the 3rd ACM conference on Electronic Commerce*, ACM (2001) 38–47
- [11] Westin, A.: *Privacy and Freedom*. (1967)
- [12] Müller, G., Flender, C., Peters, M.: Vertrauensinfrastruktur und Privatheit als ökonomische Fragestellung. In Buchmann, J., ed.: *Internet Privacy - Eine multidisziplinäre Bestandsaufnahme/A Multidisciplinary Analysis*. Springer (2012)
- [13] Adams, A.: The implications of users' multimedia privacy perceptions on communication and information privacy policies. In: *Proceedings of Telecommunications Policy Research Conference*. (1999)
- [14] Varian, H., Farrell, J., Shapiro, C.: *The economics of information technology: An introduction*. Cambridge University Press (2004)
- [15] Eisenmann, T., Parker, G., Van Alstyne, M.: Strategies for two-sided markets. *Harvard Business Review* **84** (2006) 92
- [16] Parker, G., Van Alstyne, M.: Two-sided network effects: A theory of information product design. *Management Science* (2005) 1494–1504
- [17] Bellotti, V., Sellen, A.: Design for privacy in ubiquitous computing environments. In: *Proceedings of the Third Conference on European Conference on Computer-Supported Cooperative Work (CSCW)*, Kluwer Academic Publishers (1993) 77–92
- [18] Hinz, O., Eckert, J.: The impact of search and recommendation systems on sales in electronic commerce. *Business & Information Systems Engineering* **2** (2010) 67–77
- [19] Böhme, R.: Conformity or diversity: Social implications of transparency in personal data processing. *Managing Information Risk and the Economics of Security* (2009) 291–312
- [20] Hildebrandt, M.: Profiling into the future. *Future of Identity in the Information Society (FIDIS)* **1** (2007) 1–20