

Partial Commitment – “Try Before You Buy” and “Buyer’s Remorse” for Personal Data in Big Data & Machine Learning

Lothar Fritsch^(✉)

Karlstad University, Karlstad, Sweden
lothar.fritsch@kau.se

Abstract. The concept of partial commitment is discussed in the context of personal privacy management in data science. Uncommitted, promiscuous or partially committed user’s data may either have a negative impact on model or data quality, or it may impose higher privacy compliance cost on data service providers. Many Big Data (BD) and Machine Learning (ML) scenarios involve the collection and processing of large volumes of person-related data. Data is gathered about many individuals as well as about many parameters in individuals. ML and BD both spend considerable resources on model building, learning, and data handling. It is therefore important to any BD/ML system that the input data trained and processed is of high quality, represents the use case, and is legally processed in the system. Additional cost is imposed by data protection regulation with transparency, revocation and correction rights for data subjects. Data subjects may, for several reasons, only partially accept a privacy policy, and chose to opt out, request data deletion or revoke their consent for data processing. This article discusses the concept of partial commitment and its possible applications from both the data subject and the data controller perspective in Big Data and Machine Learning.

Keywords: Big Data · Machine Learning · Data sharing · Personal information · Information privacy · Commitment · Consent · Data processing · User interface · Interaction

1 Introduction

Collection and processing of personal data is an important component of contemporary IT services. Many contemporary services are free of financial charge for end users, however they demand collection of personal data and the provisioning of advertising services as compensation. A new emerging business model for free-of-charge services is the accumulation, elaboration, analysis and selling of data provided by the users. The handling of personal data is regulated according to data protection legislation. In Europe’s General Data Protection regulation (GDPR) [1], data processors shall collect legally valid informed consent from the data subjects before they collect and process their personal data. Such informed consent should specify the scope of data collection, provide details about storage and processing, specify the purpose of data use, and indicate other parties that will get access to the data. Users are usually presented with a

privacy policy text in prose which they will have to accept and confirm as it is. Privacy policies are known to misinform [2], and to impose a high burden of responsibility on the data subjects [3]. Automatic negotiation of privacy policy references has been explored with P3P and EPAL, however is rarely found in existing systems [4]. The provision of consent is therefore, in practice, YES-NO binary decision. Service providers fulfill their legal obligation, while data subjects usually skip reading the privacy policy on their way to access the free-of-charge service. Many reasons for such behavior are found – lack of time, lack of legal understanding, pseudonymous use of services with fake identities, and non-commitment, for example for the purpose of testing the service. Data subjects might, therefore, be unaware of or ignorant about the nature of data collection and processing the service relies upon. They might accept a privacy policy with a “maybe” intention, just to proceed into using the service.

The collection of data from non-committed data subjects may, however, pose a risk to the intentions of the service provider. Dependent on the purpose of data collection, the provisioning of fake identities, incomplete or fabricated data or data patterns created through playful testing of a service may reduce the quality of the collected data. In addition, the accumulation of non-committed data subject’s data into a sample that shall represent the user population may misrepresent users upon opt-out of the uncommitted users. Non-commitment poses therefore a hazard for data quality, may endanger training data sets, statistical norm data sets, and may cause long-stranding data protection compliance obligations with respect to data protection enquiries and transparency rights.

As a solution to this problem, we suggest the introduction of partial commitment into the handling of data processing consent. We propose to extend the YES-NO choice offered today by a MAYBE option that expresses partial commitment. The remainder of the article will elaborate the background of partial commitment, discuss particular benefits both data subjects and data processors might receive from partial commitment, and drafts a research agenda for the further investigation of partial commitment to personal data processing.

1.1 Background

Commitment, or the lack thereof, has been the subject of research in many disciplines. This section reviews the results of literature research for the concept of partial commitment, delayed commitment, non-commitment and promiscuous commitment. Examples from the technology domain are the reachability manager for mobile communications which contains numerous options for policies for personal reachability for direct communications [5]. Another variant is a customer self-care interface for location services in mobile networks where customers can control fine-grained opt-in and opt-out functions against any third-party service provider [6]. One base technology for partial commitment is a reference storage for various policies which can then be, under the commitment process, referenced by the negotiating stakeholders [7]. Commitment has been discussed in the areas of risk acceptance, choice and decision-making. In psychology, a known phenomenon is a preference for the status quo. Human beings seem, when confronted with decision-making, show a preference for the status quo [8].

Reasons for this are uncertainty, incomplete information, loss aversion, complexity of the alternatives and many other aspects discussed in literature. Recent research on choice architecture deepens insight into how information presentation supports decision-making [9]. Another influential aspect of commitment is fairness in interaction. Procedural justice may improve user cooperation and data quality, as found in [10]. In addition, procedural fairness is found to increase trust in on-line applications [11]. From a trust management perspective, trust partial commitment can be assumed an integral part of pessimistic and investigative trust-building strategies [12]. A connection between privacy policies and the level of customer loyalty has been observed in recurring consumer studies on web portals [13]. Consequently, giving consent to the processing of personal data can be seen as a dialogue, not a monologue over the particularities of releasing personal data and engaging into a contract with a service provider [14]. Lack of information may cause decision procrastination in search for more information [15]. From this perspective, the usability of privacy policies can be decisive for data subject commitment, as they are part of end-user decision making [16]. There is evidence about a tight binding between good stakeholder relationship and commitment. Customer relationship management is concerned strongly with customer commitment. The importance of commitment in relationship marketing was described in [17] as: “Commitment is an important variable in the relationship marketing goal system. It is a prerequisite for the customer to proactively seek relationship maintenance whereas uncommitted customers can only be kept in relationships through instruments such as use of power, long-term contracts or in monopoly situations.”

1.2 Challenges

Many users of internet services who accept service terms & conditions and the related privacy policies are not committed at the time they sign up. They test the service, and may resign or opt out a short time in the future. Such leaving customers’ data may cause a number of issues in BD/ML systems:

- According to upcoming European data protection legislation [1], data subjects will have extensive rights concerning data protection inquiries, data export and data deletion requests from 2018 on. A BD/ML operator will have to prepare all data processing systems to comply with such requests, even for uncommitted short-term users of the services. This will cause major liabilities and compliance efforts.
- Machine Learning models trained with data gathered from non-committed data subjects may not make as good decisions as those trained with committed data subjects’ data. Service providers may be interested in separation of data acquired from committed and non-committed users. Uncommitted data subjects may “pollute” the data pool and the models.
- “Roll-back” of learning models or data collections that collect aggregated data in the case of data subject opt-out may be difficult performed on simple data bases. A roll-back mechanism for ML and for various forms for BD data aggregation should support opt-out of data subjects, including their contribution to the models and databases. Roll-back may prove useful when trying to fight pollution of models and data sets by uncommitted data subjects.

- Resulting models and databases should provide sufficient audit information about personal data processed into them, and how it contributed to model building and decision-making. Quality insurance and demonstrability of correct data processing might be essential once analysis results are questioned.

The handling of the aforementioned challenges requires strategies and techniques to handle them in an application processing data from uncommitted data subjects. In the following section, we suggest and investigate the concept of partial commitment, and how its conceptualization as a classification tools could be used to solve the challenges above.

2 Partial Commitment as a Concept: The MAYBE Button

In this section, the concept of partial commitment into processing of personal data is presented. The concept of partial commitment was suggested by Elena Barrantes for the rump session on the 11th IFIP Summer School on Privacy and Identity Management in Karlstad, Sweden, in August 2015. Lothar Fritsch moderated the discussion following the presentation. The participants – researchers, industry participants and PhD students – brainstormed about the concept, its interpretation and its uses.

The suggestion starting the brainstorming was the question whether there should be a “MAYBE button” next to the accept/decline choices when providing consent to a privacy policy (see Fig. 1). In the following sections, we will discuss the stakeholder perspectives on partial commitment. We focus on the two stakeholders “data subjects” (delivering data, expected to accept a privacy policy to access a service) and “service provider” (a personal data consuming service that expects a data subject to give some form of consent to data processing. On the rump session workshop at the 11th IFIP Summer School on Privacy and Identity Management, the participants were asked to brainstorm possible beneficial uses and implications of a “Maybe” option on privacy policies, both for data subjects and for service providers. The results were collected, analyzed and used to formulate benefits from both stakeholder groups’ perspectives, which are summarized in the following two sections.



Fig. 1. Partial commitment through the MAYBE option.

2.1 Data Subjects’ Perspective

On the rump session workshop, the participants produced four different data subject perspectives on partial commitment.

First: Why should one commit at all? Concerns were raised about how realistic a policy reflects actual data processing, how much a – yet unknown – service is worth the commitment, and about how little trust information is known about the service provider. Participants partially voiced a strong wish of ownership over their data, and voiced concerns about granting too many privileges to service providers. It was stated that there is no time to read and comprehend privacy policies, which should get compensated by possibly committing later.

Second: Inappropriateness of the privacy policy. Participants expressed concern over the appropriateness, fairness, or truthfulness of the presented privacy policy. They voiced usefulness of delayed or partial commitment where confronted with policies that are either incomprehensible (too complicated, too long, poorly written), unfair (too general, one-sided, too much power transferred to the service provider), poorly specified (written for another legal system) or technically unusable (display on devices not suitable for reading).

Third: Promiscuity - Exploration and experimentation. Participants expressed the usefulness of unconditional, playful trial options and exploration of new services. In addition, they stated that they want to be able to use several services without much consideration about the implications of their privacy policies in intersection.

Fourth: Counteraction and retaliation when faced with no choices. Participants expressed that they, in cases where they find privacy policies unacceptable, but where they have to use the services for some reason, chose obfuscation or sabotage strategies such as entering fake identities, fake data, and the intentional provocation of false profiles. The possibility of partial commitment could reduce the need for such strategies.

From the data subject’s perspective, a partial commitment can implement three different modes of interaction with a data-consuming service:

- **Promiscuity against yet unknown services or providers.** In this mode, the data subject has principal objections against commitment to a service provider. Why give exclusive rights over data and possible profits generated with it to a single stakeholder one has not yet established a relationship to, or built up trust in? Data subjects may wish to “sell” their data to several stakeholders, and chose how their data gets used freely. Depending on choices they get offered, they may delay commitment as they are not yet convinced that they have found the one service provider that suits best for their needs and requirements.
- **Test-before-commitment.** In this mode, a data subject executes the “try before you buy” philosophy. Reasons may be the satisfaction of curiosity, simple playful exploration of new services without serious commitment intentions, or mistrust in the quality of delivered service. “Try before you buy” schemes are implemented in various areas of life. In consumer protection law, when buying at the door, via telephone or on the internet, buyers can leave the contract for a certain period. Commercial providers of subscriptions, ranging from newspapers to telecommunication services,

often offer discounted trial subscriptions for limited time periods to get customers to try out new products or services.

- **Verify realities behind privacy promises.** Often, the privacy policies and service descriptions are incomprehensible to data subjects. It is hard to evaluate the implications, consequences and accuracy of privacy policies [18] and their technical and administrative enforcement [4, 19]. Data subjects may use partial commitment for the purpose of exploring and evaluation of the reality of personal data processing in the service.

The presented modes of partial commitment may help data subjects therefore help with trust establishment, help with the playful exploration and adaption of new services, and can establish a dialogue between data subjects and service providers about privacy preferences.

2.2 Service Providers' Perspective

On the rump session workshop, the participants produced four different service provider perspectives on partial commitment.

First: Measurement of privacy policy reception by data subjects. Delayed commitment could be used as a signal for poor readability or unacceptable privacy policies. Various forms of signals could help to understand customer objections. As a hypothesis, the measurement of frequencies of partial commitment was suggested: The more “maybe” commitments, the more confused or hesitant are the data subjects.

Second: Isolation of data from committed and little/not committed users. Using partial commitment, data processing services can manage separate pools of data, dependent on levels of commitment. Participants suggested that varying levels of data quality, service usage intensity and motivation of providing personal data will have a measureable impact on data quality and service quality.

Third: Focus on data consumption for Big Data applications and training sets for ML. Participants voiced concern over the accuracy of forecasting applications, ML based decisions and BD analytics when based on a data set that contains data from uncommitted or partially committed data sets. Separate data sets and models were suggested.

Fourth: Provision commitment metadata that enable rollback end reduces data management cost. Participants expected that, through available metadata on commitment levels, all forms of data management obligations (quality insurance, privacy transparency request handling, proof of foundations of automated decision-making) could be supported effectively.

From the service provider perspective, partial commitment can implement therefore three different benefits:

- **Measure the quality of privacy policies.** By assessing frequencies and detail aspects of various offered forms of partial commitment, service providers can assess the end user perspective on their privacy policies. A measurement resulting in low acceptance could then initiate a process with the aim to remove the problem.

This can be seen as the start of a communication and negotiation process for a more acceptable, and hence more customer-friendly service.

- **Separate data into classes of commitment.** Partial commitment can help with data separation along several dimensions. It can help keeping committed and uncommitted data pools separate, and may thereby improve the quality of data analysis, machine learning data sets, and decision-making. Commitment metadata may help with the deployment of services with better target population match, and may help improving the overall quality of data sets.
- **Prevent future separation and management cost.** Through suitable data classification, separation and labeling, the assessment of BD/ML decisions can better get planned, investigated, rolled back, or proven to 3rd parties. Compliance issues such as transparency and data deletion (data protection) and fairness (consumer law) can get managed better, with higher precision, and improved audibility. Systematic documentation and consideration of commitment levels may therefore prevent future cost.

In summary, partial commitment can be a tool for service providers to assess the acceptance of their privacy policies. It can be used as a tool for data separation and quality insurance, and it could, in addition, get deployed as a strategy for cost reduction, service quality improvement, and better transparency in analytics and automated decision-making.

3 Research Opportunities

From the above observation, I propose the scientific examination of the value of partial commitment in research activities. We propose to:

- Develop interaction patterns and architecture patterns for partial commitment;
- Map stakeholder needs and priorities;
- Perform usability research on user interface for partial commitment;
- Build a model for dynamic privacy management and data management with changing user commitment;
- Evaluate a prototypical implementation.

Additional interdisciplinary research opportunities can be included with:

- Research on the legal foundations, constraints and opportunities of partial commitment, e.g. through the construction of an analog to remorse periods in e-commerce or test subscriptions in telecommunications and Pay TV;
- Research on psychological aspects of usability and trust establishment between data collectors and data subjects;
- Information systems research on the influence of partial commitment on technology acceptance, diffusion, business model alignment, customer satisfaction, customer engagement, data crowdsourcing, and ad-hoc consent to data processing.

Both theoretical and applied research opportunities can be realized. In particular industry partners in the areas of Big Data, Machine Learning, Smart and autonomous

networks cars, mobile telecommunications, Internet of Things, electronic health services and marketing and customer management services should be interested in the opportunities provided by partial commitment.

4 Conclusion

I introduced the concept of partial commitment to the collection and processing of personal data. We analyzed the data subject and data processor perspective on partial commitment, followed by an identification of stakeholder benefits, including possible acceptance and trust increasing effects on the customer relationship in business models based on personal data. We showed the foundations of the concept in scientific literature, and identified a research agenda that will investigate the concept of partial commitment in the context of information privacy and data protection further, both in theory and in applied research.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), European UNION L 119/1 (2016)
2. Antón, A.I., Earp, J.B., Reese, A.: Analyzing website privacy requirements using a privacy goal taxonomy. Presented at the Proceedings of the 10th Anniversary IEEE Joint International Conference on Requirements Engineering (2002)
3. Vila, T., Greenstadt, R., Molnar, D.: Why we can't be bothered to read privacy policies: models of privacy economics as a lemons market, pp. 403–407. ACM Press, Pittsburgh (2003)
4. Stufflebeam, W.H., Antón, A.I., He, Q., Jain, N.: Specifying privacy policies with P3P and EPAL: lessons learned. Presented at the Proceedings of the 2004 ACM workshop on Privacy in the electronic society, Washington DC, USA (2004)
5. Reichenbach, M., Damker, H., Federrath, H., Rannenberg, K.: Individual management of personal reachability in mobile communication. In: Yngström, L., Carlsen, J. (eds.) *Information Security in Research and Business*. IFIP, pp. 164–174. Springer, Boston (1997). doi:[10.1007/978-0-387-35259-6_14](https://doi.org/10.1007/978-0-387-35259-6_14)
6. Zibuschka, J., Fritsch, L., Radmacher, M., Scherner, T., Rannenberg, K.: Privacy-friendly LBS: a prototype-supported case study. In: 13th Americas Conference on Information Systems (AMCIS), Keystone, Colorado, USA (2007)
7. Jøsang, A., Fritsch, L., Mahler, T.: Privacy policy referencing. In: Katsikas, S., Lopez, J., Soriano, M. (eds.) *TrustBus 2010*. LNCS, vol. 6264, pp. 129–140. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-15152-1_12](https://doi.org/10.1007/978-3-642-15152-1_12)
8. Samuelson, W., Zeckhauser, R.: Status quo bias in decision making. *J. Risk Uncertainty* **1**, 7–59 (1988)
9. Münscher, R., Vetter, M., Scheuerle, T.: A review and taxonomy of choice architecture techniques. *J. Behav. Decis. Making* **29**, 511–524 (2016)

10. Muthoo, A.: A bargaining model based on the commitment tactic. *J. Econ. Theor.* **69**, 134–152 (1996)
11. Lauer, T.W., Deng, X.: Building online trust through privacy practices. *Int. J. Inf. Secur.* **6**, 323–331 (2007)
12. Fritsch, L., Groven, A.-K., Schulz, T.: On the internet of things, trust is relative. In: Wichert, R., Laerhoven, K., Gelissen, J. (eds.) *AMI 2011. CCIS*, vol. 277, pp. 267–273. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-31479-7_46](https://doi.org/10.1007/978-3-642-31479-7_46)
13. Flavián, C., Guinalú, M.: Consumer trust, perceived security and privacy policy: three basic elements of loyalty to a web site. *Ind. Manag. Data Syst.* **106**, 601–620 (2006)
14. Coles-Kemp, L., Kani-Zabihi, E.: On-line privacy and consent: a dialogue, not a monologue. Presented at the Proceedings of the 2010 workshop on New security paradigms, Concord, Massachusetts, USA (2010)
15. Ferrari, J.R., Dovidio, J.F.: Examining behavioral processes in indecision: decisional procrastination and decision-making style. *J. Res. Pers.* **34**, 127–137 (2000)
16. Jensen, C., Potts, C.: Privacy policies as decision-making tools: an evaluation of online privacy notices. Presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria (2004)
17. Ivens, B.S., Pardo, C.: Are key account relationships different? Empirical results on supplier strategies and customer reactions. *Ind. Mark. Manag.* **36**, 470–482 (2007)
18. Sunyaev, A., Dehling, T., Taylor, P.L., Mandl, K.D.: Availability and quality of mobile health app privacy policies. *J. Am. Med. Inf. Assoc.* **22**, 1–4 (2014)
19. Earp, J.B., Antón, A.I., Aiman-Smith, L., Stufflebeam, W.H.: Examining internet privacy policies within the context of user privacy values. *IEEE Trans. Eng. Manag.* **52**, 227–237 (2005)