Chapter 19
Privacy Aspects of Recommender Systems

Arik Friedman, Bart P. Knijnenburg, Kris Vanhecke, Luc Martens, and Shlomo Berkovsky

19.1 Introduction

The deluge of online products, services, and information has made recommender systems an inherent part of the Web realm. They are used in a variety of use cases and applications: from eCommerce sites, through the Social Web, to health mobile apps. The benefits of personalized recommendations, both for users and service providers, are numerous. However, they also bring to the fore some risks that may limit the uptake of recommenders, one of which is the risk of a privacy breach.

The privacy risk is mainly caused by the recommenders' need to collect and store personal information about their users. Indeed, in order to provide personalized recommendations, a recommender needs to possess some information about its users, encapsulated in user models. This information serves as the basis for generating the recommendations and, generally, the quality of the recommendations is correlated with the amount, richness, and freshness of the underlying user modeling data. On the other hand, the same factors drive the severity of the privacy risk and

The author contributed to this chapter while he was at the University of California, Irvine.

A. Friedman (✉)
NICTA, Sydney, NSW, Australia
e-mail: arik.friedman@nicta.com.au

B.P. Knijnenburg
Clemson University, Clemson, SC, USA
e-mail: bartk@clemson.edu

K. Vanhecke • L. Martens
iMinds - Ghent University, Ghent, Belgium
e-mail: kris.vanhecke@intec.ugent.be; luc.martens@intec.ugent.be

S. Berkovsky
CSIRO, Sydney, NSW, Australia
e-mail: shlomo.berkovsky@csiro.au

F. Ricci et al. (eds.), *Recommender Systems Handbook*,
DOI 10.1007/978-1-4899-7637-6_19

the damage that can be caused if the user modeling data is exposed to third parties. This is referred to as the *privacy-personalization trade-off* [10, 24, 37, 87, 98, 156], and it inevitably manifests once personalized recommendations are considered.

The privacy risks posed by personalization are aggravated when more sophisticated recommendation scenarios are deployed. For example, consider a recommender that, as part of the recommendation process, either augments its user models by extracting new features and populating their data, or cross-links multiple sources of user modeling data. In these scenarios, the recommender is likely to uncover additional information that was not readily accessible in the original user models, i.e., information that the users may not have consented to be released for the recommendation purposes. Having this information exposed and accessed by untrusted parties could lead to harmful consequences.

In this chapter we concentrate on the privacy challenge faced by recommender systems. We survey related work on privacy-enhanced recommenders and partition it into three broad categories. The first focuses on *architectures* that facilitate more private recommendations. These entail various decentralized solutions that eliminate a single repository of user modeling data, which would otherwise be the target for attacks on the recommender. The second category refers to *algorithmic* solutions, which either perturb the original user modeling data or apply formal encryption methods. These assure that, even if accessed by an untrusted party, only modified/encrypted user data would be exposed, rather than the original data. Lastly, the category of *policy* driven solutions addresses directives and legislation initiatives that limit the storage, transfer, and exploitation of personal user data. Clearly, these solutions are not mutually exclusive, and a recommender may—and often will—deploy solutions from multiple categories.

While these solutions may improve objective and measurable privacy aspects, an important question pertains to the users of the recommenders. They may have their own considerations regarding the sensitivity of their data, exposure/preservation of some information, and measures they are willing to take to protect their privacy [21]. Hence, users' perception of and reasoning about privacy deserves special attention. Therefore, we also discuss users' privacy attitudes and behaviors, as well as current practices and recent advances to support the users' privacy decision-making process.

This chapter is structured as follows. In Sect. 19.2 we give a broad definition of privacy and discuss the privacy risks faced by the users of recommender systems. In Sect. 19.3 we outline the three categories of solutions to these risks; namely, the architectural (Sect. 19.3.1), algorithmic (Sect. 19.3.2), and policy solutions (Sect. 19.3.3). We survey a number of papers implementing the solutions and summarize each category. In Sect. 19.4 we switch to the human aspects, and discuss users' perception of and attitude towards privacy, as well as privacy-related decision making. We conclude the chapter in Sect. 19.5, where we outline the achievements and shortcomings of privacy-enhanced recommender systems and discuss future research directions in light of emerging trends in recommendation technologies.

19.2 Privacy Risks in Recommender Systems

Most scholars argue that in the modern information age people regard their personal information as a *commodity*: they are willing to give up some personal information in return for personal gains. Recommender systems are a perfect example of this dynamic: They collect a wide variety of user data as input for their recommender systems, and in return provide their users with better services and products [10, 37, 44, 87, 139]. The information collected might include users' clicking or viewing behavior; contextual information like the location or mood; social information like user friends, family, or colleagues; as well as demographic parameters like age and occupation [72]. To make sure that data collectors treat the collected information responsibly, the OECD [114] has defined a set of Fair Information Practices (FIPS):

Collection Limitation Data should be collected within limits, by lawful and fair means and with consent (where appropriate).

Data Quality Data should be relevant, accurate, complete and kept up-to-date.

Purpose Specification The purposes of collection should be specified at the time of collection.

Use Limitation Data should not be used or disclosed for other purposes except with consent or by the authority of law.

Security Safeguards Personal data should be protected against unauthorized access, destruction, use, modification or disclosure.

Openness Users should be able to know what data is being collected, who controls the data, and for what purposes they are used.

Individual Participation An individual should be allowed to inspect the collected data about themselves, and have them erased, rectified, completed or amended.

Accountability The collector of the data should be accountable for complying with the above measures.

Generally speaking, privacy is breached when any of these principles are violated. Given their need to collect large amounts of information and innate capability to infer users' personal tastes from this data, recommender systems run a heightened risk to violate the Collection Limitation, Purpose Specification, Use Limitation, and Security Safeguards principles. In this light, we categorize privacy risks in Table 19.1 along two dimensions: whether the privacy breach is due to direct access to existing data (a violation of the Collection and Use Limitation principles) or due to inference of new data (a violation of the Purpose Specification principle), and who the adversary trying to uncover user information is. We consider three types of adversaries: (1) the *recommender system* interacts with the user, but it might operate in a way that is incompatible with the user's expectations of privacy (a violation of the Collection and Use Limitation principles); (2) *other users* of the system have no direct access to another user's private data, but they might exploit the outputs of the recommender to uncover the information of a target user (a violation of the Security Safeguards principle); and finally, (3) *external entities* are not users

Table 19.1 Privacy risks in recommender systems

Adversary	Direct access to existing data	Inference of new data
Recommender system	Unsolicited data collection	Exposure of sensitive information
	Sharing data with third parties	Targeted advertising
	Unsolicited access by employees	Discrimination
Other users	Leaks through shared device or service	Inference from the recommender output
External entities	Lawful data disclosure	Exposure of sensitive information
	Hacking	
	Re-identification of anonymized data	

of the recommender, but they may try to access the information retained by the system or intervene in the interaction between the system and its users to get access to such information (another violation of the Security Safeguards principle, but regarding a different type of security safeguard). We next look in detail at the risks imposed by each of these actors.

19.2.1 Risks Imposed by the Recommender System

19.2.1.1 Direct Access to Data

Recommender systems typically rely on a central entity, which accesses personal user data for the purpose of personalizing a service. However, the availability of this information, combined with commercial incentives, may result in this data being used in a way that violates the end-users' expectations of privacy, even when this use is consistent with the provider's privacy policy [46]. There are several ways in which direct access to data could expose users to privacy risks, including:

Unsolicited data collection As storage capabilities are cheap, online services are tempted to collect as much user data as possible, either because it might be useful at some point in the future (e.g., Chap. 6 discusses the value of rich contextual information), or because it can be monetized. However, collection of data that is not deemed necessary to provide a service may break user expectations of privacy. For example, in a survey that aimed to capture the expectations of what sensitive resources mobile apps use [99], Pandora Internet Radio was one of the apps singled out by the users for unexpected resource usage, since it accesses the contact list on the mobile device. In general, users seem particularly wary of "context tracking," arguably because unwanted or unexpected inferences can be made about such data [80].

Sharing data with third parties There are many scenarios in which recommender systems have incentives to share raw user data with third parties. For example:

- Companies that have access to such information may wish to share it to collaborate with the research community, as was the case when AOL released anonymized user search queries [12] and in the Netflix Prize competition [20].
- Companies may need to share data with third parties to outsource parts of their operation. Today, many companies offer so-called *recommendations as a service*. The third party receives user profiles and interaction logs from a website, processes them, generates recommendations, and sends them back to the website. While the user profiles may have been anonymized before transmission, a copy of the user profile now exists with the third party. Even if the user were to delete their account, they could not verify the deletion of their profile by the third party.
- Finally, service providers may be tempted to sell personal user data to data brokers, as this was shown to be a lucrative business [17]. Data may also change hands following acquisition of companies, or when liquidators sell off databases of bankrupt companies.

Ackerman et al. [1] and Krishnamurthy and Willis [92] highlighted that propagation to third parties and profile data that can be linked back to a user's identity are important concerns that users have when they consider releasing information online. Although the data custodian may take precautions and anonymize the data prior to release to safeguard user privacy, the released data may be subject to de-anonymization attacks, as will be discussed later.

Unsolicited access by employees While the recommender system may take precautions to ensure user data is maintained under its control, it is possible that employees, who need access to user data to fulfill their role, will abuse their privileges to snoop for data of people they know. Employees may also be tempted to steal the data of well-known people (celebrities) for curiosity or for money. This risk exists in any system that retains user information, and can be mitigated to some extent by ensuring appropriate access control and auditing mechanisms.

19.2.1.2 Inference from User Preference Data

Sophisticated manipulation of the data collected or processed by the recommender system (see Chap. 7 for an overview of data mining methods) could lead to additional privacy risks due to inference of new data, sometimes without the awareness or consent of the user:

Exposure of sensitive information Several recent works [36, 91, 147] have demonstrated the power of machine learning techniques in uncovering sensitive and private personal information, including personality traits (see Chap. 21). While such inferences are probabilistic in nature, they could be harmful even if wrong, particularly when judgments are based on risk (e.g., insurance decisions) or prejudice (e.g., workplace discrimination).

Targeted advertising In targeted advertising, the collected data is used to learn user interests and select advertisements that are most likely to result

in conversion. The targeted ads may expose sensitive or embarrassing information—one prominent example is of a parent who learned that his teenager daughter was pregnant after Target started sending her coupons for baby clothes and cribs [47].

Discrimination Recent works [105, 106] have shown evidence of online price discrimination facilitated by personal information. Individuals may perceive this as a misuse of their information, and as overstepping the purposes specified for data collection.

Inference attacks exploit various aspects of user data to derive sensitive and private information. These attacks typically rely on correlations learned from other users' data, but can exploit them in various ways. For example, an adversary can rely solely on information contributed by the system users [147], leverage semantic relations between different attributes [36], cross-link the data with additional sources to extract more correlations [91], or exploit the structure of social links [157].

Weinsberg et al. [147] showed that demographic information such as age, gender, ethnicity, or political orientation can be inferred from information disclosed to recommender systems. Several classifiers were trained using the data contributed by the users, and inferred with high accuracy the demographic information of users who did not disclose similar data. Experiments conducted on the Flixster and Movielens datasets demonstrated the effectiveness of the approach. In fact, the mere act of watching a movie (regardless of the rating) conveys a lot of information, in the sense that classifiers trained over binary data (i.e., movie watched or not) performed only slightly worse than those trained on the complete rating data.

While Weinsberg et al. exploited structured data, Chaabane et al. [36] leveraged the ontologized version of Wikipedia to identify semantic relations between unstructured user interests, and showed how seemingly harmless interests, such as music interests, can leak sensitive information about users. They assigned the user interests into higher-level interest topics, and the interests of each user were mapped to these topics, allowing to identify users with similar tastes. Assuming that users with similar tastes are similar in multiple aspects, it was then possible to guess a user's private attribute based on the public attributes of similar users. The authors crawled public profiles from Facebook, and used the self-declared, publicly available music interests of users to infer their gender, relationship, age, and country attributes.

Kosinski et al. [91] conducted a large-scale study that correlated the Facebook 'likes' of users to a range of sensitive personal attributes, including sexual orientation, ethnicity, religious and political views, and personality traits using machine learning techniques. The authors generated predictors for these sensitive attributes and achieved remarkable results. For example, the model could distinguish between homosexual and heterosexual men in 88 % of cases, African Americans and Caucasian Americans in 95 % of cases, and between Democrats and Republicans in 85 % of cases. While some 'likes' were related to the attribute in question (e.g., liking pages related to homosexuality), some of the discovered correlations had no obvious connections.

The inference problem is exacerbated in online social networks, where friendship links and group membership can be leveraged to infer private information.

Zheleva and Getoor [157] considered the possibility that linked objects in a social network are correlated, i.e., that online friends share common characteristics. They proposed several inference attacks that exploit the structure of the network to predict private attributes. Based on evaluation of such inference using data from Flickr, Facebook, Dogster, and BibSonomy, the authors concluded that the performance of the predictors was dataset-dependent. For example, link-based methods did not perform well, since there was no strong correlation between the inferred attributes and the friends. On the other hand, group membership improved the inference, and some of the group memberships allowed to predict the user's attributes with high accuracy. Note that while users may have control over which attributes are made public, in some social networks (e.g., Facebook and Flickr) the user has limited control over the visibility of group membership information.

19.2.2 Risks Imposed by Other System Users

Since recommender systems leverage data collected from numerous users, they allow users to learn personal information about each other, even when such information is kept private. This problem is most evident when users share the same account on a device or a service: the recommendations for this account would be derived from the users' combined activities, and therefore the recommendations generated for one user provide insights on the activities of the other users. A similar problem can occur in group recommender systems (see Chap. 22).

A harder problem is imposed when the outputs of a recommender system leak private information of other unrelated users in the system. This problem is particular to collaborative filtering recommender systems (as opposed to content-based recommenders), since inherently these recommenders adapt the recommendations provided to each user based on data collected from other users. Ramakrishnan et al. [124] showed how the recommendations and their explanations can expose information of users who rate items across disparate domains. The recommendations allow an adversary to deduce connections between items. For example, given a certain item, an adversary can create a fake account and add item ratings to identify the smallest set of items that would result in a recommendation of the target item. This implies that there exists a set of users who rated both these items and the target item. This set of users is likely to be small when the items belong to different domains, making it easier to target these users in privacy attacks. For example, the revealed connections can be combined with additional data sources to compromise the identity of the users and uncover additional personal information.

A stronger attack that exploits the public outputs of item-to-item collaborative filtering systems was put forward by Calandrino et al. [31]. Public outputs of such recommenders typically contain item similarity lists or cross-item correlations. For example, Amazon provides the "customers who bought this item also bought…" lists, Hunch provides the entire item-to-item covariance matrix, and Last.fm provides an item similarity list. By passively observing the changes in

these outputs over time, an attacker could infer private transactions of a target user, given background knowledge on some items previously rated by the user. In an item-to-item collaborative recommender, when a user makes a transaction involving an item, this results in an increase of the similarity of the item to other items in the user's transaction history. Therefore, the attacker can track the similarity lists of items known to be associated with the target user, and identify new items in the lists. When the same item appears in a number of tracked lists, the attacker can infer that the item was added to the target user's record. The authors successfully applied this approach to several real-world recommender systems, including Hunch, LibraryThing, Last.fm and Amazon. The attack exhibits a trade-off between the number of inferences and their accuracy (for example, inference results on LibraryThing ranged from 58 inferences per user with 50 % accuracy to six inferences per user with 90 % accuracy) and achieves the best results when applied to small or new sites.

In addition to the passive attack that Calandrino et al. presented in [31], they also described an active sybil attack that targets neighborhood-based collaborative filtering. Given background knowledge on some items previously rated by a user, the adversary creates fake users that are similar to the target user, and likely to be identified as neighbors of that user and of each other. A neighborhood-based recommender is therefore likely to provide to the fake users recommendations based on the other fake users and the target user. This allows to isolate the target user's data, as any recommended item that does not appear in the fake profiles is likely to originate from the target user.

19.2.3 Risks Imposed by External Entities

Data sharing and misuse are subject to the control of the recommender system, and may therefore be mitigated through regulation, or be disclosed to obtain the user's consent. In contrast, some scenarios may lead to unintended data disclosure. One risk is imposed by unlawful access to data by hackers (e.g., due to insufficient security safeguards), resulting in data theft. Another risk is due to court subpoenas and surveillance by law enforcement agencies. While such data access is lawful, it is often conducted without user awareness, and, in some cases, even without the service provider's awareness.

Third parties may also obtain personal information gathered by recommender systems after it was anonymized for privacy protection. However, even in the anonymized form, this data poses a serious privacy risk due to the possibility of de-anonymization. Narayanan and Shmatikov [108] demonstrated the difficulty of guaranteeing anonymity in transaction and preference records common in recommender systems. In general, the sparseness of large multi-dimensional data collections ensures that a record will not have many other "similar" records in the dataset, allowing to single it out and re-identify it with relatively little background information. The attack can be carried out by an adversary who knows

a (possibly imprecise) subset of the target user's attributes, e.g., items that were rated by the user, ratings that were assigned, or the time of the ratings. The de-anonymization algorithms evaluate the similarity of each record in the anonymized dataset to the background information. Due to the sparseness of transaction and preference records, these algorithms are robust to imprecision and uncertainty in the background knowledge, as well as to a moderate level of perturbation in the published records. The authors conjectured that the amount of perturbation needed to defeat this de-anonymization approach would destroy the utility for collaborative filtering.

The effectiveness of this attack was demonstrated using the Netflix Prize dataset, containing anonymized ratings of 500K Netflix subscribers. The authors found that with background knowledge consisting of eight movie ratings (of which two may be wrong) and rating dates known within a 14-day error, 99 % of records can be uniquely re-identified in the dataset. Even without knowing the dates on which the items were rated, information about a few rated items may be sufficient. For example, 84 % of records can be uniquely re-identified if the adversary knows six out of eight movies rated outside the 500 most frequently rated movies. This background information may be relatively easy to obtain for most users, e.g., by observing their voluntary disclosure of information on social networks or on IMDB. It can be argued that the anonymized records may not contain sensitive data. However, even in these cases, re-identification carries a privacy risk: any information that can be traced back to a person can be leveraged in subsequent attacks, and provide additional hooks that the adversary could use to de-anonymize further data releases. An aggregate of such releases could lead to a "database of ruin" [115], which would tie together digital traces from different sources, exposing an elaborate picture on individuals' online and offline activities.

The possibility of re-identification of the Netflix dataset resulted in a lawsuit that was settled out of court, and subsequent cancelation of the second Netflix challenge [29]. To date, safe release of de-anonymized datasets for research purposes is still an open problem. As stated in [108], in such scenarios "*the purpose of the data release is to foster computations on the data that have not even been foreseen at the time of release, and are more sophisticated than the computations that we know how to perform in a privacy-preserving manner*." Inferences on this data, thus, pose privacy problems, because they almost definitely go beyond users' initial expectations of privacy.

19.2.4 Summary

Research conducted in recent years demonstrated the ability to infer highly sensitive information from user interest data, even when they express seemingly innocent information. Such information could be abused either by the systems that collect the data (e.g., inferring users' psychological traits and leveraging these for targeted

advertising); by other users in the systems who may be exposed to the data (e.g., by default, public "likes" on Facebook) or may analyze the output of the recommender; or by external entities that access the user data.

These results stress that even privacy-conscious users who may withhold some of their information, cannot guarantee their privacy, since the withheld information could be inferred from other information disclosed to the recommender. Moreover, the privacy of a user does not solely depend on the user's personal choices and privacy preferences, but is also influenced by the data made available by other users, regardless of whether they are associated with the user. Therefore, the user may only have limited control over the privacy risks resulting from using the system. Instead, integrating privacy into the design of recommender systems may prove more effective in safeguarding users' privacy. In the next section we will discuss approaches that can be taken to mitigate the identified privacy risks.

19.3 Privacy Solutions

The discussion about the risks of personal data leakage through recommender systems naturally leads to the "defender" side, i.e., how can the recommender protect user privacy without compromising the quality of the recommendations. We consider three categories of approaches, which can address the privacy problem in recommender systems:

- The first category refers to *architectures*, *platforms*, and *standards* that minimize the data leakage threat. These include various protocols and certificates that guarantee to users that the recommendation provider adheres to privacy-preserving practices and protects the users' personal data with due diligence. This inherently limits the ability of external entities to access user data or to infer new data, other than the authorized and regulated data access methods. We classify into this category also the distributed architectures, which eliminate the single point of failure typical to centralized recommenders.
- The second deals with the *algorithmic techniques* for data protection. Here, we distinguish between several types of approaches. Some of them involve data modification approaches—either of user identities (identity anonymization or abstraction to stereotypes) or of the rating data (substituting or adding noise to true rating data). Others exploit provable privacy guarantees offered by the differential privacy framework or apply cryptographic tools to protect the data. The basic idea underpinning the algorithmic techniques is that even if the users' personal data leaked to an adversary or untrusted party, they would possess only modified or encrypted information, and would struggle to recover the original data.
- The third category refers to "top-down" *legislations*, *policies*, and *regulations*, which may be imposed on the recommendation services by their governments and legislative bodies, or adopted as self-regulatory industry practices. They may

preclude the services from manipulating, sharing, or trading the data. Although this category of approaches addresses outright many of the above privacy risks, the regulations vary significantly across countries and even states, and their enforcement is hard to validate in practice.

The main rationale for this categorization lies in the grouping of these three categories into technical and non-technical solutions. The former consist of the architectural and algorithmic solutions, whereas the latter includes only the policy solutions. The technical solutions either provide a general infrastructure that supports privacy, or offer specific algorithms for data protection. On the other hand, the non-technical solutions provide an umbrella that outlines the allowed and the prohibited activities with regards to personal user data. Another important observation stemming from this grouping is that although the three categories seem independent, many recommender systems may (and actually should) apply more than one approach to protect the privacy of their users. Hence, we propose recommender system designers to consider all three categories of solutions when devising their privacy-protection mechanisms.

For example, consider a use case of a large-scale eCommerce website providing personalized recommendations to users. The site may apply architectural solutions and distribute the data storage. At the same time, the site may exploit algorithmic techniques and allow only cryptography-protected data access. In addition, the site may want to increase user trust and declare that the collection and use of personal user data is done in compliance with privacy regulations. Many of these details, especially the architectural and the algorithmic solutions in place, are not disclosed by practical websites. Nevertheless, we refer the reader to several publicly accessible privacy policies (see those of eBay,[1] Amazon,[2] and Google.[3])

We would like to revisit the access and inference risks outlined in Table 19.1, and intersect these with the three categories of solutions. Clearly, the architectural and policy solutions better address the direct data access risk, as private protocols, distribution of the recommendation process, and data protecting regulations make unauthorized access to the data harder. The application of algorithmic approaches cannot eliminate this access, but reduces the value of the data if it gets accessed. However, the algorithmic approaches substantially minimize the risk of inferring new data, as the input to the inference attacks becomes unreliable. It should also be mentioned that the policy solutions are likely to address the data inference risk, as they often prohibit the use of the collected data for purposes that are beyond those declared by the data collector.

In the following sections we elaborate on each of the categories and on specific works that apply these approaches.

[1] http://pages.ebay.com/help/policies/privacy-policy.html.

[2] http://www.amazon.com/gp/help/customer/display.html?nodeId=468496.

[3] http://www.google.com/intl/en/policies/privacy/.

19.3.1 Architecture and System Design Solutions

In this section, we consider how the architecture underlying the recommender system can put hard limits on the disclosure, propagation and linkability [119] of profile data. In Sect. 19.3.1.1, we introduce a trusted component that is guaranteed to act in a certain way. Then, in Sect. 19.3.1.2, we look at an architecture for social networking websites that gives the user control over their profile data through standard technologies from the Semantic Web. Finally, in Sect. 19.3.1.3, we cover approaches that shift some of the workload of the recommender system to the client-side, thereby reducing the amount of user data that needs to be disclosed.

19.3.1.1 Trusted Software for Limiting Linkability and Propagation of User Data

As we saw in Sect. 19.2, a recommender system may cross-link data from multiple sources to create comprehensive user models. If the models are retained after the recommendation process terminates, or even disclosed to untrusted parties, this could pose a grave threat to the user's privacy. The recommender could therefore make certain claims regarding data storage, linkability, and disclosure, to put the user's mind at ease, e.g., "no disclosure of any profile data without explicit consent," "no linkability between individual user sessions," "no linkability between partial user profiles," or "temporal limits on the storage of user data."

But how can the user trust that the service actually complies with these principles? In researching privacy-preserving recommendation solutions, Cissée and Albayrak [39] identified three ways of establishing trust:

- Reputation [74]: Non-compliance would lead to negative user feedback and sentiment, which discourages other users from using the service.
- Certification [136]: A trusted third party performs a detailed technical audit, e.g., by analyzing the source code and performing tests, to verify that the software has all the qualities and properties that it claims to have.
- Trusted computing [56]: An application has the ability to verify that a system consists of specific hardware and software, e.g., the ability to encrypt data in a way that can only be decrypted in a particular configuration.

We will analyze two examples of trusted systems that restrict the linkability and propagation of profile data: a privacy-preserving event planner proposed in [39] and a privacy-friendly loyalty card and shopping assistant application for smartphones.

In [39], Cissée and Albayrak built a privacy-preserving event planner on top of a FIPA-compliant [137] multi-agent system (MAS). The authors list various properties of MAS entities that make them ideal for creating a privacy-preserving recommender system, in which only trusted parties can temporarily cross-link user profile data from multiple sources: entities are autonomous and can be deployed dynamically in the MAS environment; each entity can perform a well-defined task; entities can communicate with each other; and they can be tamper resistant.

With regard to user privacy, the purpose of the system is to ensure that disclosed user profile data is not stored permanently and cannot be linked to any particular user. A temporary filter agent (TFE), responsible for generating recommendations, is created, and a relay entity establishes control over the TFE's communication abilities on the user's behalf. This way, it can be ensured that only the recommendations will be propagated to other entities; user profile data will not be propagated because the relay does not provide the TFE with the means of communicating it to other parties. Controlling agents' communication abilities is not part of the standard MAS feature set, so the authors have implemented this aspect as trusted software. With control established, the user provides profile data (made up of behavior information, personal details and preferences) to the TFE and the service provider hands the TFE a set of items to recommend from. The TFE uses all data at its disposal to generate content-based recommendations for the user, which are then propagated to the service provider for visualization. Finally, the TFE is terminated by the relay entity, thereby destroying the linked dataset. The service provider can thus present the user with personalized recommendations without gaining permanent access to the profile data.

The MobCom project[4] explored the possibility of implementing various identity-based applications such as identity cards, membership cards, and customer loyalty cards on a smartphone, in a way that protects the privacy of the user. Put et al. [123] developed a shopping and loyalty card application that discloses only the minimal amount of information required, with user consent. The smartphone serves as a self-scanning device with secure local storage for the customer's personal information, shopping history, loyalty points and product vouchers. At the start of each shopping session, a temporary shopping basket is created under a new pseudonym, so that the store cannot track customer behavior across sessions. In exchange for disclosing profile data, e.g., product preferences, the retail store offers a more personalized service and additional loyalty points. This way, customers control their data and can weigh the benefits of releasing profile data against the loss of privacy. In this architecture, both the smartphone application and the in-store service are regarded as trusted software. At the start of the shopping session, the smartphone and the server can verify that each is running the trusted software and that it has not been tampered with. The smartphone does not release any profile data without the user's explicit consent. The shopping basket contents and any disclosed profile data are destroyed at the end of the session.

19.3.1.2 User-Managed Portable Profiles

Beyond the privacy risks originating from inference and profiling, which were discussed in Sect. 19.2, social networking websites (see also Chap. 15) tend to become data silos [27], with profile data either locked away or only partially

[4] http://www.mobcom.org.

accessible through proprietary APIs. If users were able to port profile data from one platform to another, they could receive better recommendations and more personalized service, alleviating the cold start issue when joining a new service. They could also allow access to specific profile information on a case by case basis. Currently, however, this scenario is not possible because users do not have such level of control over their data.

We focus here on an alternative architecture proposed by Heitmann et al. [64], which puts the user in charge of fully portable profile data through Semantic Web technologies and an access control system. Using this architecture, profile data can be shared between services and the users can decide what parts of their profiles are disclosed to each provider. Building on earlier work of Hollenbach et al. [66], Heitmann et al. base their architecture on three standards: (1) Friend-of-a-Friend [26]: a data format suitable for storing generalized user profile data, as well as social friendship relations; (2) WebID [32]: an SSL certificate that refers to the URI where the profile data can be found; and (3) a Web Access Control [66]: vocabulary for controlling access rights to resources. The authors also identify three distinct roles for entities that wish to participate in the architecture:

- *Profile stores* are tasked with storing the user profiles and providing access to data according to the access rules. They also allow users to manage these access rules. Notably, the user can perform this role by hosting his own profile.
- *Data consumers* are third-party services that wish to access the user profile data. Each time they request data from a profile store, data consumers authenticate themselves with their own unique WebID.
- *User agents* are responsible for authenticating the user with profile stores and data consumers through their WebIDs.

To summarize, users are able to port their profile data from one service to another. By using Semantic Web technologies, entities that wish to perform any of these roles, have an easy-to-use, stable, and non-proprietary interface to work with. Users can selectively disclose parts of their profiles to data consumers of their choice. Through the use of WebIDs, unlinkability of data is built-in: a user can have multiple identities, each with its own WebID. Data consumers are thus unable to link multiple WebIDs to a particular user and the framework assumes that the profile stores can be trusted to not maintain or disclose links between the users' multiple identities. We refer to Chap. 4 for more on Semantic Web technologies.

19.3.1.3 Generating Recommendations on the Client

Shifting some of the recommender's load to client devices allows to reduce the amount of information accessed and retained by a recommendation service, thereby mitigating any privacy risks that could result from the server's exposure to user data.

Several works proposed to implement the recommendation process as a pure peer-to-peer system, thereby eliminating the role of a centralized service [22, 94]. However, such systems could still expose user data to other users, who now interact

directly with the user to generate the recommendations. Lathia et al. [94] addressed this risk by proposing a privacy-friendly measure of similarity that relies on the *concordance* between users, i.e., the proportion by which two user rating sets agree. This measure has the property that it can be evaluated by comparing the two sets of ratings to a third rating set, rather than directly to each other. Therefore, user similarity can be evaluated without exchanging user profiles. Berkovsky et al. [22] leveraged a hierarchical topology, in which peers are organized into peer-groups managed by super-peers. A user who seeks recommendations interacts with the super-peers. The super-peers select a random subset of the underlying peers, aggregate the results obtained from them, and return them to the querying user, who processes them to generate the recommendation.

In a hybrid approach proposed by Shokri et al. [131], each client interacts with a centralized server to obtain recommendations, but can also exchange information with other system users to enhance privacy. In this approach, each user maintains two profiles: an offline profile stored locally at the client, which is updated continuously, and an online profile at the server that is only synchronized occasionally. Users contact each other and exchange items, so that their offline and consequently the online profiles are a mix of each user's original ratings and ratings provided by other users. To maintain accurate recommendations, the exchange process favors ratings conducted by similar users.

One of the challenges in distributed architectures is that many recommendation algorithms are computationally intensive, and while mobile devices have recently become powerful, they are still ill-suited for heavy computations. This limitation gives rise to architectural approaches that divide work between a powerful back-end and a weaker end-user device. Such approaches allow for recommendations to be generated on the client, while disclosing less information to the centralized recommender back-end than in a centralized recommendation scenario. These approaches usually leverage the ability to break the recommendation generation into two stages: (1) modeling, for which the entire dataset is typically required, and (2) recommending, for which the models are used to compute the recommendations. Given an established model, recommending can be a relatively light-weight task.

For example, consider item-based collaborative filtering, where all the available user-item ratings are needed to construct the item-to-item similarity matrix. Recommendations are then generated by taking items that are similar to items that the user has previously consumed. In PocketLens [107], Miller et al. set out to build a portable collaborative filtering recommender system, where the similarity computation is separated from the recommendation stage. Through homomorphic encryption methods that are also applied in secure voting systems, the back-end constructs an item-to-item similarity matrix based on co-occurrence, without having to decrypt individual purchase records. A mobile client can retrieve this matrix and generate recommendations locally. After implementing and evaluating several architectures, the authors found that their best performing architecture could protect the user's privacy without compromising the recommendation accuracy.

The separation between the modeling and the recommendation stages is also evident in matrix factorization. The modeling stage that consists of the derivation of

the latent factors, requires access to all the ratings and is computationally expensive. The recommendation generation then is realized as a product of two latent vectors and can be performed on the client. Moreover, since matrix factorization separates between the user and the item latent factors, the user data can be stored on the client side. Vallet et al. [141] explored this possibility in a semi-decentralized setting, in which the server maintains item factors, whereas user factors are stored and maintained on the client-side. The authors developed a streaming model, which performs incremental updates of the latent factors using only the data of the user interacting with the system, and without any server-side retention of user data. The predictive accuracy of this model was found comparable to that of a system that retains user data.

Isaacman et al. [70] leverage the same matrix factorization property in the context of a distributed system of content producers (e.g., bloggers) and consumers. To maintain privacy, information is exchanged only between the content producer and its subscribers, e.g., item ratings are shared only with the item's producer. The system computes the probability distribution of content ratings that is estimated with a low-rank latent model constructed by solving the factorization problem. Each producer maintains a factor vector that constitutes its "production profile." In addition, each consumer maintains for each possible rating value a factor vector, and these factor vectors constitute its "consumption profile." The client can compute the product of these vectors to estimate the probability that the consumer would provide a certain rating to any given producer's content, without disclosing all of the consumer's ratings to that producer.

To summarize, architectures that shift computation to the client side are particularly useful for mitigating privacy risks that follow from data retention on a centralized server. However, user data may still be exposed during the interaction with the server, or when interacting with other system users. Cryptographic protocols allow to address this deficiency, and are discussed in detail in Sect. 19.3.2.4.

19.3.2 Algorithmic Solutions

In this section, we discuss algorithmic solutions to recommender system privacy. We split them into four categories: algorithms based on pseudonyms or user anonymization, algorithms involving user data modification, differentially private algorithms, and cryptography-based algorithms. Similarly to what was discussed earlier, these categories are not mutually exclusive; a recommender may benefit from employing multiple solutions that belong to different categories.

19.3.2.1 Pseudonyms and Anonymization

Algorithmic approaches that mask the users of recommender systems through pseudonyms and anonymization were not received well initially. In particular, Schafer et al. [130] wrote in 2001 that "*anonymizing techniques are disasters for recommenders, because they make it impossible for the recommender to easily recognize the customer, limiting the ability even to collect data, much less to make accurate recommendations*." More than a decade later, the topic still remains largely under-investigated and there are only several works in this direction.

An early proposal for a pseudonymity-based personalization framework was developed by Arlein et al. [9] and drew on the notion of 'personae.' The framework implied that users have in place a suite of abstractions of themselves, e.g., entertainment, medical, and shopping, and use these abstract entities when interacting with various websites and services. Each persona is linked across multiple services and exposes only the activities carried out by this persona. The services access only one user persona at a time and cannot link it to other personae, so they are unable to uncover additional information, while the users manage their own personae and set access rights for various services and abstractions.

Another pseudonimity framework for personalized systems was proposed by Kobsa and Schreck [89]. The framework includes a suite of privacy-preserving components: user anonymization, user data encryption, role-based access, and selective access permissions. Each component is managed by a dedicated server and the servers tune the overall level of user privacy to the user's privacy settings and the degree of cooperation between the services possessing the partial user models.

The approaches to user anonymization in recommender systems typically entail simple de-identification solutions. For example, in the Netflix Prize data, the identities of the users were replaced with random numbers. A major threat to this anonymization method lies in the high dimensionality and sparsity of the data [108], which is typical in recommender datasets. As discussed in Sect. 19.2.3, this sparsity can be exploited to thwart anonymization and re-identify the records.

19.3.2.2 Obfuscation

Application of data perturbation (or obfuscation) techniques to recommender systems was inspired by earlier works outside the field of recommender systems [7]. The basic idea underpinning this body of work is that modifying a certain number of data points in the user profiles, e.g., by adding noise to the real data, will have a limited effect on the recommendation accuracy. However, if adversaries or an untrusted party accessed the user profiles, they would only obtain the disguised profiles. This allows for "plausible deniability" [61, 142]: the adversary cannot prove whether a certain profile entry is accurate.

To the best of our knowledge, this idea was first proposed for recommender systems by Polat and Du [120]. They used a randomized data perturbation technique to mask ratings stored in the user profiles. The data is modified by adding random

noise to the ratings, such that no certain information about the ratings can be derived. Since the recommendations are generated by aggregating user ratings, the overall impact of data perturbation on the recommendations is assumed to be minor. The authors compared the recommendations generated using the masked data with those using the original data, and showed that perturbed profiles could still generate reasonably accurate recommendations. The accuracy of the recommendations is inversely correlated with the magnitude of the noise, but the impact of noise decreases with the number of users and items accessible by the recommender.

Another variant of data perturbation was presented by Parameswaran and Bloug [118]. They proposed to mask auxiliary data pertaining either to users (e.g., demographic data) or to items (e.g., domain metadata), which are exploited by the similarity computation mechanism of collaborative filtering. The evaluation showed that the impact of masking auxiliary data on the accuracy of the recommendations is minor, although the direct contribution of this perturbation to user privacy was not explored.

Unfortunately, data perturbation through the addition of noise is inapplicable to binary data, which is prevalent in recommenders, as the systems increasingly rely on binary behavior logs (browsing logs, purchase data, listened songs, etc.). In this case, the addition of noise distorts the logs and can be easily identified. In [122], Polat and Du applied a different technique, called a randomized response, to the binary user profiles. This technique randomly chooses which bits of the binary profile are preserved and which are flipped. Two variants of randomized responses were evaluated and, as before, the accuracy was found to be correlated with the volume of training data.

The application of random perturbation has gone beyond the canonic collaborative filtering. Yakut and Polat [153] applied data perturbation also to the Eigenstate-based variant of CF that reduces the dimensionality of the rating matrix through Principal Component Analysis. Two distributions for generating the noise factors and several variants of privacy-enhanced Eigenstate CF were proposed and evaluated. Also, Kaleli and Polat [76] applied randomized response to a Naïve Bayes Classifier implementation of CF. That work primarily focused on tweaking the noise parameters for the purpose of maintaining reasonable levels of user privacy and recommendation accuracy at the same time.

Basu et al. [14] applied data perturbation to the Slope-One recommender [96], a highly scalable version of item-based collaborative filtering. It was found that Slope-One is robust to the noise and capable of delivering reasonable accurate recommendations despite the masking of user data. Polat and Du [121] applied data perturbation to an SVD-based CF recommender, which decomposes the masked ratings matrix into a product of three latent matrices. SVD recommendations were also found to be reasonably robust to random perturbation.

More recently, data perturbation was applied by Renckes et al. [126] to a hybrid graph-based recommender representing users as nodes and their similarity through the edges. The paper reaffirmed the findings of Polat and Du [120] relating to the impact of data availability on the accuracy of private recommendations, and practically demonstrated the privacy-accuracy trade-off. In a nutshell, privacy loss

decreased with the level of perturbation, but the accuracy of the recommendations deteriorated too, such that privacy and accuracy conflicted with each other. To allow users more control over the privacy-accuracy trade-off, Kandappu et al. [77] have proposed an interactive obfuscation mechanism. The obfuscation is applied to ratings before they are shared with the system (input perturbation). Before sharing new ratings, the mechanism probes the recommender to obtain rating predictions over a hold-out set of items, which were rated by the user but were not disclosed to the recommender. The magnitude of obfuscation is then calibrated based on the accuracy of those predictions, such that privacy protection is maximized within the constraints of a target accuracy level.

Berkovsky et al. [23, 24] focused on the application of data perturbation to various ratings in collaborative profiles. They compared the impact of five data masking policies applied to both *moderate* (close to average) and *extreme* (positive or negative) ratings on the accuracy of the generated recommendations. Perturbation of the latter was found to have a higher impact on the accuracy of the recommendations than of the former. That is, extreme ratings bear more information than moderate ratings, and adding noise to these ratings deteriorates the accuracy of the recommendations. However, extreme ratings were perceived as more sensitive by the users. This gives a different perspective on the privacy-accuracy trade-off, as masking the sensitive ratings damages the recommendation accuracy.

Aside from a potential decrease in recommendation accuracy, data perturbation can also be problematic for legal and psychological reasons. A perturbed profile is essentially "incorrect data," which violates the Data Quality principle of the FIPS (see Sect. 19.2) as well as several European privacy laws that require data collectors to pursue the correctness of the collected data. Psychologically speaking, users may fear that this incorrect data may result in incorrect inferences (which is possible in specific instances even when the overall accuracy of the recommender does not decrease due to perturbation). Even worse, if users' data gets subpoenaed or stolen and published, they may have a hard time defending the claim that some of the data in their profile is incorrect. So while obfuscated data may afford users "plausible deniability," it does not offer them what we would like to call "deniable plausibility" (i.e., the ability to prove that certain items were in fact fabricated by the obfuscation mechanism). Indeed, a study by Chen et al. [38] on the application of obfuscation techniques in online social networks has indicated that users care about the impact of obfuscation on their visible profile, and suggested to incorporate such preferences into the obfuscation algorithms.

We summarize the surveyed works that apply data obfuscation techniques in Table 19.2. These are split into the basic Collaborative Filtering (based on either user-to-user or item-to-item similarity) and other CF algorithms.

19.3.2.3 Differential Privacy

Differential privacy [48] is a privacy model based on the principle that the output of a computation should not allow inference about any record in the input. This is achieved by requiring that the probability distribution over the possible

Table 19.2 Privacy-preserving recommendation algorithms with data obfuscation

Similarity-based collaborative filtering (CF)	User-to-user similarity [24, 120]
	Item-to-item similarity [118, 122]
Other CF algorithms	Eigenstate-based CF [153]
	Naïve Bayes CF [76]
	Slope-one [14]
	SVD-based CF [121]
	Graph-based recommender [126]

outcomes does not change significantly when any particular record is added to or removed from the input. Therefore, differential privacy provides the means to mitigate inference of private user data from the output of the recommender system. One of the commonly used approaches to obtain differential privacy is through the Laplace mechanism, in which carefully calibrated noise sampled from the Laplace distribution is added to a computation. The noise masks the influence that any difference in a particular record could have on the outcome of the computation.

McSherry and Mironov were the first to study the application of differential privacy to recommender systems, and in particular to collaborative filtering [103]. They used the Laplace mechanism to derive noisy counts and sums over the input ratings, and to compute a differentially-private variant of the item-to-item covariance matrix. The noisy covariance matrix could then be used to generate differentially-private k-Nearest Neighbors and SVD recommendations.

Zhu et al. [158] took a different approach to differentially private neighborhood-based collaborative recommendations, aiming specifically at the sybil attack presented by Calandrino et al. [31] (see Sect. 19.2.2). They considered a differentially-private k-nearest neighbors algorithm that operates in two steps: selection of the neighbors, and rating prediction based on the neighbors. They relied on the smooth sensitivity [112] of the similarity function, allowing to introduce lower levels of noise than those required by the Laplace mechanism. They also introduced randomness to the k nearest neighbors selection, while ensuring that, with high probability, the selected neighbors have high similarity scores.

Machanavajjhala et al. [101] studied privacy-preserving social recommendations on the basis of a graph linking users and items. Given the graph, they derived utility vectors that capture the utility of items for users, with the goal of inducing a probability distribution that maximizes the user's utility while keeping the utility vector private. The authors provided a theoretical analysis of the problem and concluded that good recommendations were achievable only under weak privacy parameters, or only for a small fraction of users, highlighting that the privacy-accuracy trade-off also exists in differential privacy based methods.

Riboni and Bettini [127] investigated the application of differential privacy to context-aware recommendations, and specifically to recommendations of Points of Interest (POI), where the spatial context is taken into account. The spatial domain of the service is partitioned into non-overlapping regions, and each POI belongs to a

single region. In addition, each user belongs to a given stereotype, which represents semantic abstraction of profile data. The Laplace mechanism is used to capture the distribution of POI preferences for each stereotype. Consequently, when a user queries a region, the POIs best matching the user stereotype are recommended.

The research of differentially private recommender systems shows that while in some settings (e.g., social recommendations) it may be impossible to obtain privacy and accuracy guarantees simultaneously, in other cases privacy-preserving recommender systems can achieve reasonable accuracy. However, the works conducted so far assume a one-off computation, whereas re-calculation of recommendations when additional data becomes available may introduce additional privacy leaks. Therefore, maintaining privacy over multiple computations or data releases requires an increase in the amount of introduced noise, and leads to deterioration in accuracy. While there is a line of work studying efficient differential privacy in continual settings [49, 57], this has not been studied yet in recommender systems.

19.3.2.4 Cryptographic Solutions

Cryptographic solutions mitigate privacy risks triggered by the exposure of user data, like intentional misuse (e.g., sharing data with third parties or inferring sensitive information), as well as unintentional disclosure (e.g., data theft). Secure multi-party computation protocols allow to accurately compute recommendations, while keeping user input confidential. Unlike data obfuscation or differential privacy, secure computations produce the same recommendations as non-private protocols, but this comes at the cost of computational overhead, making these protocols suitable mainly for off-line recommendations.

The majority of the work in this area relies on additive homomorphic encryption schemes, such as the Paillier public-key cryptosystem [116]. Essentially, in such encryption schemes, any linear function of the inputs can be evaluated by manipulating their encryptions. This property has been leveraged in several recommendation algorithms and architectures, listed in Table 19.3. Below, we elaborate on the proposed architectures and provide examples of homomorphic encryption applications.

Distributed settings As detailed in Sect. 19.3.1.3, distributed architectures mitigate privacy risks by keeping the data on the client side. To the best of our knowledge, the protocol proposed by Canny [32] was the first application of secure multi-party computations to recommender systems. A partial singular value decomposition of the ratings data can be reduced to a series of additions of user inputs and carried out over encrypted inputs using an additive homomorphic encryption. Based on this, Canny proposed a peer-to-peer system, consisting of two types of nodes: "clients" who provide in each iteration their encrypted contribution to the gradient, and "talliers" who manipulate and aggregate these inputs to derive an encrypted total gradient. The encryption key is shared between the clients, and each client applies its share of the key to decrypt the total.

Table 19.3 Privacy-preserving recommendation algorithms with homomorphic encryption

Distributed	Weighted slope-one [13]
	Neighborhood-based [51]
	Trust networks [65]
	Partial SVD [32]
	Factor analysis model [33]
Cross-system collaboration	User-to-user similarity [71]
	Item-to-item similarity [154]
Client-server	Weighted slope-one in cloud setting [15]
Privacy service provider	General framework [8]
	Neighborhood-based [53]
	Trust networks [52]

If enough clients provide decryptions with their share of the key, then the talliers can reconstruct the new gradient. The result of the computation is guaranteed to be correct, even in the presence of malicious parties, as long as a sufficient portion of the nodes are trustworthy and follow the protocol.

Cross-system collaboration Distributed algorithms can also be carried out between service providers, allowing cross-system collaboration without disclosing clients' information to other systems, and thereby mitigating privacy risks due to sharing data with third parties. For example, Jeckmans et al. [71] studied how a company can generate recommendations based on its own customer data and data from other companies, while keeping customer data confidential. They relied on additive homomorphic encryption, as well as secure comparison, absolute value, and division protocols. The proposed two-party protocol, executed between a pair of servers, allows to generate predictions based on user-to-user similarity, which is evaluated using the ratings that the users have on both sites.

Client-server settings Encryption can keep user ratings confidential when the user interacts with the server in the prediction stage, as demonstrated in a Slope-One recommender that Basu et al. [15] studied. In a Slope-One predictor, predictions are based on the average deviations of item ratings, which are linear combinations of user ratings, making it suitable for secure evaluation with additive homomorphic encryption. In the learning phase, the users send their (obfuscated or anonymized) inputs to the cloud in the clear, and the cloud application produces the deviation matrix and the cardinality matrix for the Slope-One predictor. In the prediction stage, the target user sends a rating vector encrypted with a public key, which the cloud application manipulates with additive homomorphic encryption to produce an encrypted prediction vector. Finally, the user decrypts the vector to retrieve the prediction.

Privacy service provider Several works addressed the privacy risks in the client-server interaction by introducing a third party acting as a privacy service provider.

These solutions rely on the "division of trust" principle [8], i.e., no entity in the system holds the complete information. Aïmeur et al. [8] proposed a framework for privacy preserving recommenders based on this principle. Each merchant in the system is assigned to an agent that mediates the interaction with the clients. The client profiles are encrypted with the agent's public key, such that the agent can access them but the merchant cannot. On the other hand, the items are anonymized by a mapping known only to the merchant, so the agent cannot know the actual products purchased or rated by the customer. The agent maintains the list of products associated with a cluster of clients and a table of product similarities, uses these to generate recommendations, and can update them based on user inputs, but without knowing the actual products.

Homomorphic encryption is not the only approach to secure computation of recommendations. Nikolaenko et al. [109] proposed a privacy-preserving matrix factorization algorithm, in which the recommender profiles items without learning the users' ratings. In the proposed protocol, the recommender is assisted by a crypto-service provider, who prepares a Yao garbled circuit [155] that evaluates the item profiles given the encrypted rating inputs. The authors report a reasonably low running time and, since the described operations are parallelizable, they suggest that the algorithm may be suitable for batch processing of real large-scale datasets.

The extensive research on cryptographic solutions for privacy-preserving recommendations shows the feasibility of these solutions in diverse settings and with different recommendation algorithms. However, these solutions entail significant computational resources and time, as well as storage and communication overhead, which still impose a hurdle for their application in online recommender systems.

19.3.3 Policy Solutions

As Kobsa points out [87], many countries and states actively regulate consumers' privacy, and many industries adopt additional privacy guidelines. We refer to [144] for an overview of the impact of privacy laws and regulations on personalized systems up to 2006. Two important proposals since then are the U.S. Consumer Privacy Bill of Rights [67] and the 2012 revision of the European Privacy Directive [55].

Both of these proposals have a heavy emphasis on transparency and control. For example, the U.S. Consumer Privacy Bill of Rights suggests that "*companies should offer consumers clear and simple choices [...] about personal data collection, use, and disclosure*" and "*companies should provide clear descriptions of [...] why they need the data, how they will use it*" [67]. Under the European Privacy Directive, "personal data should be processed on the basis of the consent of the person concerned or some other legitimate basis" [55].

The U.S. privacy bill furthermore requires that consumers are able to access the personal data that companies collect about them, and correct it if necessary. It also requires that data collection is focused and limited to what is expected in the context

in which the data was provided by the consumer. The European directive also requires that people are able to access their personal data. It additionally requires that they are allowed to transfer this data from one service to another, and that people are able to delete their data should they so desire.

The 2002 version of the European Union Privacy directive severely limited the use of non-essential cookies, often used for personalized advertising [54]. As a result, online advertising could not be targeted and became far less effective in the EU than in other countries [60]. The new directive requires websites to explicitly ask their users to accept its non-essential cookies. The Netherlands and the United Kingdom [69] have already implemented this directive as a national "cookie consent" law. However, to comply with the rules without losing advertising money, most sites give users only two options: leave the website or accept the cookies and continue. The resulting sprawl of consent-requesting pop-ups has caused much confusion among users, who typically accept the cookies without knowing what they really consent to, which arguably only increases their privacy concerns [140].

An alternative to privacy legislation is self-regulation via trust seals like the TRUSTe seal [19] or privacy standards like P3P [43]. Xu et al. [151] have shown that TRUSTe seals can be an effective substitute to legislation when it comes to reducing consumers' privacy concerns. TRUSTe seals have been shown to reduce perceived risk and increase trust, whereas P3P compliance increases trust but does not reduce perceived risk [150]. Self-regulation is not without problems, though. Research has shown that trust seals are only partially effective [50, 68, 128], and A/B tests on eCommerce websites have demonstrated that seals may lead to significantly lower conversion rates [30, 59]. This calls the benefits of "certification" (cf. [136]) into question. P3P, on the other hand, suffers from poor observability and complex user agents, which has led to a low level of adoption on the user-side [16].

In conclusion, privacy legislation and regulation has become more comprehensive over the last few decades. However, as Compañò and Lusoli point out, "*policy makers need to take into account that citizens do not always behave rationally*" [40], a topic we will cover in much more detail in the next section.

19.4 Human Aspects and Perception of Privacy

While we have mainly discussed the technical solutions to privacy risks in recommender systems, the concept of privacy is an inherently human attitude associated with the collection, distribution and use of disclosed data, and this disclosure is also a human behavior. Since recommenders critically rely on their users to disclose information about themselves, recommender system developers are advised to conduct user experiments to study users' information disclosure behavior and their privacy-related attitudes towards the recommender system (see Chap. 9).

This section discusses existing research concerning users' privacy attitudes and behaviors. The link between privacy attitudes and subsequent behaviors is not very clear: while several studies find this link to be significant [80, 88, 132] others

find that it is not, or at least not very strong [2, 4, 58]. Due to this divergence, which Norberg et al. call the *privacy paradox* [113], developers of recommender systems are advised to study users' attitudes *and* behaviors regarding the privacy of their systems. The privacy paradox is a symptom of the fact that users' cognitive resources are in most cases insufficient to effectively take control over their privacy. The end of this section therefore discusses the importance of supporting users to make better privacy decisions, as well as an interesting new venue for recommender systems to provide such "privacy decision support."

Privacy Attitudes In studying privacy attitudes, one can make a distinction between privacy attitude as a personal trait or tendency, and as an attitude directed towards a specific system. General privacy concern was first measured by Westin and Harris and Associates, who classified people into three categories: privacy fundamentalists, pragmatists, and unconcerned [63, 148]. Researchers have since recognized that this personal trait consists of multiple dimensions. For example, the Concern For Information Privacy scale consists of four correlated factors: collection concerns, unauthorized access, fear of accidental errors, and secondary use [133]. Similarly, Malhotra et al. provide an Internet Users Information Privacy Concern scale measuring three factors: collection, control, and awareness [102].

Several works have highlighted the importance of measuring privacy concerns as a system/context-specific concept [6, 18, 132]. System-specific factors considered in previous work include "perceived privacy threats" [80, 88, 149], "perceived protection" [88], and "trust in the company" [80, 104]. These system-specific factors are usually better at predicting users' disclosure behavior than privacy concerns as a personal trait. Recommender system developers are thus advised to measure users' system-specific privacy attitudes. Moreover, they should not just focus on protecting users' privacy via the technical means described earlier in this chapter, but also to reduce the potential privacy threats to begin with (a philosophy called "privacy by design", cf. [34]) or to increase the reputation of their brand.

Privacy Behaviors Laufer and Wolfe were the first to argue that people trade off the risks and benefits of disclosure [95], a process that Culnan and Bies have called "privacy calculus" [45]. This term is commonly used to investigate information disclosure [62, 97, 149], and has become a well-established concept in privacy research [132]. In the field of recommender systems, several researchers have demonstrated that users indeed make this trade-off when deciding what information to disclose [10, 37, 58, 80, 85, 88, 90, 98]. The exact outcome of this trade-off depends on the context of the decision [81, 110]. Particularly, if users deem the requested information relevant to the purpose of the system, they will be more likely to disclose it. For example, it is reasonable to expect that a system for recommending nearby restaurants would collect street-level location information from the user device, but a user may be surprised to learn about such data being collected by a book recommender. This can be problematic for recommender systems, since they often use data from diverse application domains.

19.4.1 The Limits of Transparency and Control

Having a minimum level of control over one's disclosure is a necessary prerequisite for being able to engage in a privacy calculus. Moreover, people can only make an informed trade-off between benefits and risks if they are given adequate information. Based on this reasoning, advocates of transparency and control argue that they empower users to regulate their privacy at the desired level [35, 138, 152]. This advocacy for transparency and control has become a central part of the privacy directives proposed in the European Union and the United States [55, 67].

The call for control suggests that recommender systems should provide users advanced capabilities to manage their privacy. However, while users *claim* to want full control over their data, they typically eschew the hassle of actually exploiting this control [40]. While it is possible to overcome this control paradox [81], the privacy controls of systems like Facebook are so complex that they are overwhelming or confusing to most users [42]. As a result, Facebook users have severe misconceptions about the implications of their selected privacy settings [100].

Similarly, the call for transparency suggests that recommender systems should be forced to be open about their privacy practices, so that users can walk away if they do not like them (cf. "reputation" [74]). However, Bakos et al. demonstrate that only 0.2 % of all users read boilerplate documents such as End User License Agreements [11]. As noted earlier, "summarizing" this information with trust seals may actually impede rather than increase system usage [30, 59].

This ironic effect of trust seals on privacy concerns extends to other privacy-related situations as well. For example, John et al. demonstrate that even subtle privacy-minded designs and information may trigger users' privacy fears and reduce disclosure and participation [73]. They found that a professional looking site garners higher privacy concerns than an informal and unprofessional looking site, because the former design reminds users of privacy. While it is arguably more risky to entrust such an unprofessional-looking site with one's information, its appearance apparently downplays privacy concerns and increases disclosure.

Arguably, since even a professional looking site can instill privacy concerns, any reference to privacy will inadvertently prime users with privacy fears. This highlights a fundamental problem of any privacy-preserving architecture or algorithm: informing users about the superior privacy protection is likely to make them more concerned about their privacy [78, 80]. In some cases, this fear stems from concerns that the developers of these systems had not accounted for. For example, Kobsa et al. show that while client-side recommendation algorithms prevent the disclosure of personal information to third parties, users are concerned about their device getting lost or stolen [88]. Their user profile could then not only fall in the hands of a third party; they themselves would lose access to it. Users' lack of familiarity with a technology may exacerbate their privacy concerns. For example, Kobsa et al. show that users are rather skeptical about cloud-based recommendation services [88] like those proposed in [15].

The proponents of increasing transparency and control in information disclosure decisions assume that people are rational decision-makers who will use the provided information and controls to their best advantage. However, our decisions often do not follow rational economic principles [75] (see also Chap. 18), and this also holds true for information disclosure decisions [4, 5]. In fact, information disclosure decisions are among the hardest decisions to make, because they have delayed and uncertain repercussions that are difficult to trade-off with the possibly immediate gratification of disclosure [2, 5]. In this light, an abundance of information and control may only aggravate this problem, because it can lead to choice overload or information overload. Consequently, several researchers have recently questioned the effectiveness of the "transparency and control" paradigm [111, 135].

19.4.2 Privacy Nudges

The first step in supporting users' privacy decisions that does not require users to be rational decision-makers is to nudge these decisions into the "right direction" [3, 146] (see below for a discussion regarding what the "right direction" of privacy nudges could be). A nudge is a subtle yet persuasive cue that makes people more likely to decide in one direction or the other. Carefully designed nudges make it easier for people to make the right choice, without limiting their ability to choose freely. Broadly speaking, two types of nudges have been tried out in the field of privacy decision-making: justifications and defaults.

Justifications Justifications make it easier to rationalize decisions, and to minimize the regret associated with choosing the wrong option. Different types of justifications include providing a reason for requesting the information [41], highlighting the benefits of disclosure [90, 143], and appealing to the social norm [6, 25]. Justifications are especially useful in recommender systems, because recommenders are able to extract valuable taste information from seemingly irrelevant data. A good disclosure justification can nudge users to disclose these data, which helps to build their user model and improve the accuracy of the recommendations.

The effect of justifications seems to vary though. In a study by Kobsa and Teltzrow, users were 8.3 % more likely to disclose information when they knew the benefits of disclosure [90]. In a study by Acquisti et al. users were 27 % more likely to do this when they learned that many others decided to disclose the same information [6]. However, Besmer et al. found that social cues had barely any effect on users' Facebook privacy settings: only the small subset of users who take the time to customize their settings may be influenced by strong negative social cues [25]. Knijnenburg et al. tested a wide range of justifications in a demographics- and context-based mobile app recommender [80, 84]. They also found "fickleness" in the effects of justifications on users' disclosure to—and satisfaction with—the recommender. Users found these justifications helpful, but

in contrast to some of the above findings, the justifications did not increase users' disclosure, trust, or satisfaction with the system, but rather decreased them. In line with Besmer et al. [25], Knijnenburg and Kobsa conclude in a follow-up analysis that only a subset of users is amenable to justifications [79].

Defaults The other approach to nudging users' privacy decisions is to ease their burden of making information disclosure decisions by providing sensible defaults (see Chap. 18). Providing a certain default option may nudge users in the direction of that default. For example, John et al. [73] show that people are more likely to admit to certain sensitive behaviors via an act of omission than via an act of commission. Similarly, Lai and Hui [93] show that defaults have a significant impact on user participation in an online newsletter. Recommender systems can manage privacy perceptions by carefully setting the defaults of optional features such as making one's taste profile public, or social network integration.

Another default that can be used to nudge privacy decisions is the order of the disclosure requests. Acquisti et al. demonstrated that people disclose less information when requests are made in increasing order of intrusiveness compared to a random order [6]. This effect is particularly pronounced for more intrusive questions: asking those questions upfront increases their likelihood of being answered. Arguably, people become more wary of disclosing very personal information as the disclosed information accumulates; the most relevant information should thus be requested upfront. Similarly, Knijnenburg and Kobsa manipulated the request order, and showed that any type of information enjoys higher disclosure when requested first rather than last [80, 84]. Note though, that although asking sensitive questions upfront increases disclosure in research settings, it may scare away new users when done in commercial applications. The order of disclosure requests arguably has a large impact in conversational recommender systems, where quick convergence on an accurate user model needs to be balanced with privacy concerns related to sensitive information requests. Disclosure request order strategies are thus an important topic for future research in recommender systems.

The problem with existing privacy nudging techniques is that they have to take an implicit stance on whether the purpose of the nudge should be to increase disclosure, or to decrease it. Recommender system developers may claim that it is in users' best interest to provide more data to the recommender, as it will improve their user model and, subsequently, the recommendations. They may thus argue to use nudges to increase disclosure, but these nudges may cause the more privacy-minded users to feel "tricked" into disclosing more information than they would like [28]. Others (e.g., privacy advocates, certain lawmakers) may instead believe that privacy is an absolute right that needs to be defended at all costs. But if the protective nudges they impose make it more difficult to disclose information, this would reduce the overall benefit of a recommender system, especially for less privacy-minded users.

19.4.3 Privacy Adaptation

Given these opposing forces, how can we nudge users in the "right direction?" This is a difficult question, because human decisions are highly dependent on the personal context in which they are made, and the same holds true for information disclosure decisions [5, 73, 97, 110]. For example, the fact that one person has no problems disclosing a certain item in a particular context does not mean that disclosure is equally likely for a different person, a different item, or in a different context [82, 97]. Likewise, a convincing justification to disclose a certain item in a particular context for a certain person, may be completely irrelevant for a different person, a different item, or a different context [25, 79]. The "right direction" of a privacy nudge thus depends on these contextual variables. This idea of context-dependent privacy nudges leads to a new application domain for recommender systems: *user-tailored privacy decision support* [86, 145]. Specifically, a recommender can be used to predict users' context-dependent privacy preferences based on their known characteristics and behaviors, and then provide automatic "smart default" settings [134] in line with their disclosure profiles. Below we outline the budding research in this new field of "privacy adaptation."

The first step towards privacy adaptation is to gain a deeper understanding of people's cognitive decision-making process: What kind of benefits and threats do users consider when making disclosure decisions? What is the relative weight of each of these aspects? Can the weights be influenced by a justification or a default, and if so, in what context(s)?

Some of the work by Knijnenburg et al. tries to measure these cognitive determinants and integrate them in behavioral models of information disclosure decisions. For example, they demonstrate that:

- the effect of justifications on information disclosure decisions is mediated by the user's perceptions of *help*, *trust* and *satisfaction* [80];
- the effect of decision context in a location-sharing service depends on users' perception of the *privacy* and *benefits* of the available options [83] (so-called "context effects;" cf. Chap. 18);
- perceived *risk* and *relevance* mediate user evaluation of the purpose-specificity of information disclosure requests [81].

The second step towards privacy adaptation is to determine how information disclosure depends on the recipient, item and type of user. This would allow to train a recommender that can tailor defaults and justifications to these contextual factors. Work in this direction shows that even though privacy preferences vary considerably across users, recommendation techniques can be used to predict these preferences quite accurately. For example, Knijnenburg et al. identify distinct subgroups of users with similar privacy preferences in many domains [82]. These subgroups can be mapped to demographics and other behaviors, allowing a recommender to classify users into a certain subgroup. Ravichandran et al. [125] apply k-means clustering to users' contextualized location sharing decisions to come up with a number of

default policies. They show that a small number of default policies for the user to choose from could accurately capture a large part of their location sharing decisions. Sadeh et al. [129] apply a kNN algorithm and a random forest algorithm to learn users' privacy preferences in a location-sharing system. They show that the applied recommendation techniques can help users in specifying more accurate disclosure preferences. Pallapa et al. [117] propose context-aware approaches to privacy preservation in wireless and mobile pervasive environments. One of their solutions leverages the history of interaction between users to determine the level of privacy required in new situations. They demonstrate that this solution efficiently supports users in dealing with their privacy concerns. Finally, adaptive procedures also work for justifications: although justifications generally do not increase disclosure or satisfaction, Knijnenburg and Kobsa find that tailoring justifications to the user can reduce this negative effect [79].

In sum, privacy adaptation strikes a balance between giving users no control over, or information about, their privacy at all and giving them full control and information. It solves the problem of finding the "right direction" for nudges by using users' own preferences as a yardstick. At the same time, it gives users the right privacy-related information and the right amount of privacy control that is useful, but not overwhelming. It thereby enables users to make privacy-related decisions within the limits of their bounded rationality. In many systems, privacy concerns seem to rise in concert with the complexity of users' privacy decisions. "Privacy adaptation" may thus present a unique opportunity for recommender systems to help solving this problem.

19.5 Summary and Discussion

We conclude the chapter with a summary of the privacy-enhancing solutions that were outlined, along with their current shortcomings. Next, we discuss current and emerging trends in recommender systems and identify the key privacy issues associated with them. Lastly, we suggest research tracks to better address the privacy risks of today and those of the future.

In Sect. 19.2, we discussed various privacy risks originating from the recommender system itself, from other system users, or from third parties. The risks are highly diverse, but center around potential adversaries either directly accessing the existing user data or inferring new information through cross-linking multiple sources of user data. While they can be broadly categorized as either technical or non-technical, the solutions that were proposed in Sect. 19.3 are even more diverse than the challenges they seek to address.

The architectural solutions covered various protocols and certificates that guarantee that the recommender behaves in a way that preserves the users' privacy. Barriers are put up for untrusted parties that may want to access user profiles or infer non-disclosed sensitive data. Then, we proceeded to algorithmic solutions, which incorporate privacy into the recommendation generation process. Here, we

partitioned prior works into four broad directions: to anonymize and/or abstract individual users; to introduce noise into the original user data, making it hard to uncover true user preferences; to use differential privacy, a widely-used model that offers provable privacy guarantees; and to exploit cryptography-based approaches to generate recommendations, while keeping user inputs confidential. Note that these directions are by no means mutually exclusive—a recommender may deploy algorithms from several groups to improve user privacy.

As discussed in Sect. 19.4, users have their own perceptions of privacy that do not necessarily align with the privacy assurances provided by the above solutions. Moreover, some of the proposed approaches may even have an opposite effect on users' behavior and their perception of privacy. We suggested that privacy solutions should be tailored to users' inherent privacy preferences. This results in a new opportunity for recommender systems: providing privacy decision support.

The field of recommender systems is still largely evolving, and is gaining emerging popularity in several relatively new use-cases and application domains. Some of these applications pose a significant risk to user privacy, and, therefore, the importance of privacy-preserving recommendations is paramount there. We will briefly discuss some of these cases and highlight their privacy implications.

Recommenders on the Social Web. Online social networks are tremendously popular these days. The social Web attracts billions of users, who not only expose unprecedented volumes of personal information, but also voluntarily cross-link data from a wide spectrum of sources. Various personalization and recommendation technologies have been developed for the social Web (Chap. 15), and these highlight the need for privacy-preserving solutions that will deliver user-tailored services without compromising user privacy.

Cross-Domain Recommender Systems. The challenge of generating recommendations by combining multiple sources of user modeling data, which potentially span several recommender systems and application domains, has recently attracted a lot of attention (Chap. 27). This poses a direct threat to privacy, as domain-specific user profiles are inherently linked, and cross-domain recommenders already apply the techniques mentioned in Sect. 19.2 for the inference of new undisclosed data. Hence, cross-domain recommenders call for a special focus on the preservation of user privacy.

Mobile and Context-Aware Recommendations. Users are increasingly surrounded by sensors and smart environments, which interact directly with the users' personal devices. This facilitates the collection of rich user profiles and opens the opportunity for the delivery of context-aware recommendations (Chaps. 6 and 14). Users have little control over these pervasive data collection procedures. Users may wish to control data access limitations and the invasiveness of the recommender, but since these recommenders typically operate in the background, the act of control itself may disrupt the users' primary workflow. Control mechanisms should thus be very lightweight, lest users simply ignore them.

Explanation of Recommendations. Recommendations are often accompanied by a textual description explaining why the items were recommended to the user (Chap. 10). Consider Amazon's "you were recommended X because you bought

Y" or the widely-used "people who examined X were also interested in Y." While this helps users to find related items and supports the vendors' cross-selling, these explanations can potentially compromise user privacy by leaking private information and revealing information to others watching over the user's shoulder.

Group Recommenders. Consumption of the recommended items is increasingly used in a group setting, where users' individual preferences are combined to provide recommendations that fit the entire group (Chap. 22). In these settings, users may infer the preferences of the members of their group from their combined recommendations. Group recommender systems may thus need to perturb recommendations in a way that allows users a certain level of "plausible deniability" regarding their specific tastes and preferences.

This chapter has presented a patchwork of technical and non-technical solutions that can each address *specific* privacy risks regarding current and future recommender system scenarios. However, it should be highlighted that most of the existing works in the recommender systems space are focused on a single solution and very little has been done on developing a *holistic* and encompassing solution. Hence, the challenge of integrating the diverse (and often conflicting) solutions from the architectural and algorithmic realms, and developing a recommender that is privacy-friendly at the core, user-friendly and maintainable from a development point of view, and, not the least, complies with existing privacy policies, is still open.

At the same time, recommender system developers should not forget that dealing with privacy extends beyond the technical aspects of their systems. The privacy attitudes among recommender system users—the yardstick against which privacy practices should be evaluated—vary considerably and evolve continuously. Therefore, industry players have to engage in an active conversation with their users about what are considered good privacy practices. As we ran a quick survey among industry contacts to get a basic understanding of prevalent industry privacy practices, it became painfully clear that companies do not feel comfortable to talk about even their basic approach to privacy. Moving forward, though, we predict that a more conscientious discussion about privacy in recommender systems will emerge, and we conjecture that the key challenges presented above will be addressed both by the research community and by industrial players concerned with improving the privacy of their customers.

References

1. Ackerman, M.S., Cranor, L.F., Reagle, J.: Privacy in e-commerce: Examining user scenarios and privacy preferences. In: Proceedings of the 1st ACM Conference on Electronic Commerce, EC '99, pp. 1–8. ACM, New York, NY, USA (1999). DOI 10.1145/336992.336995
2. Acquisti, A.: Privacy in electronic commerce and the economics of immediate gratification. In: Proceedings of the 5th ACM conference on Electronic commerce, EC '04, pp. 21–29. ACM, New York, NY (2004). DOI 10.1145/988772.988777
3. Acquisti, A.: Nudging privacy: The behavioral economics of personal information. IEEE Security and Privacy **7**, 82–85 (2009). DOI http://dx.doi.org/10.1109/MSP.2009.163

4. Acquisti, A., Grossklags, J.: Privacy and rationality in individual decision making. IEEE Security & Privacy **3**(1), 26–33 (2005). DOI 10.1109/MSP.2005.22
5. Acquisti, A., Grossklags, J.: What can behavioral economics teach us about privacy? In: A. Acquisti, S. De Capitani di Vimercati, S. Gritzalis, C. Lambrinoudakis (eds.) Digital Privacy: Theory, Technologies, and Practices, pp. 363–377. Auerbach Publications (2008)
6. Acquisti, A., John, L.K., Loewenstein, G.: The impact of relative standards on the propensity to disclose. Journal of Marketing Research **49**(2), 160–174 (2012). DOI 10.1509/jmr.09.0215
7. Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: SIGMOD Conference, pp. 439–450 (2000)
8. Aïmeur, E., Brassard, G., Fernandez, J.M., Mani Onana, F.S.: Alambic: A privacy-preserving recommender system for electronic commerce. International Journal of Information Security **7**(5), 307–334 (2008). DOI 10.1007/s10207-007-0049-3
9. Arlein, R.M., Jai, B., Jakobsson, M., Monrose, F., Reiter, M.K.: Privacy-preserving global customization. In: Proceedings of the 2nd ACM Conference on Electronic Commerce, EC '00, pp. 176–184. ACM, New York, NY, USA (2000). DOI 10.1145/352871.352891
10. Awad, N.F., Krishnan, M.S.: The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. MIS Quarterly **30**(1), 13–28 (2006)
11. Bakos, Y., Marotta-Wurgler, F., Trossen, D.R.: Does anyone read the fine print? testing a law and economics approach toStandard form contracts (2009). URL http://archive.nyu.edu/handle/2451/29503
12. Barbaro, M., Zeller Jr., T.: A face is exposed for AOL searcher no. 4417749. URL http://www.nytimes.com/2006/08/09/technology/09aol.html. [Online; accessed 22-January-2014]
13. Basu, A., Kikuchi, H., Vaidya, J.: Privacy-preserving weighted Slope One predictor for Item-based Collaborative Filtering. In: Proceedings of the international workshop on Trust and Privacy in Distributed Information Processing, Copenhagen, Denmark (2011)
14. Basu, A., Vaidya, J., Kikuchi, H.: Perturbation based privacy preserving slope one predictors for collaborative filtering. In: IFIPTM, pp. 17–35 (2012)
15. Basu, A., Vaidya, J., Kikuchi, H., Dimitrakos, T.: Privacy-preserving collaborative filtering on the cloud and practical implementation experiences. In: IEEE CLOUD, pp. 406–413 (2013)
16. Beatty, P., Reay, I., Dick, S., Miller, J.: P3P adoption on e-commerce web sites: A survey and analysis. IEEE Internet Computing **11**(2), 65–71 (2007). DOI 10.1109/MIC.2007.45
17. Beckett, L.: Big data brokers: They know everything about you and sell it to the highest bidder (18 March 2013).
18. Bélanger, F., Crossler, R.E.: Privacy in the digital age: A review of information privacy research in information systems. MIS Quarterly **35**(4), 1017–1042 (2011)
19. Benassi, P.: TRUSTe: an online privacy seal program. Commun. ACM **42**(2), 56–59 (1999)
20. Bennett, J., Lanning, S.: The netflix prize. In: KDD Cup (2007)
21. Berkovsky, S., Borisov, N., Eytani, Y., Kuflik, T., Ricci, F.: Examining users' attitude towards privacy preserving collaborative filtering. Proceedings of DM. UM **7** (2007)
22. Berkovsky, S., Eytani, Y., Kuflik, T., Ricci, F.: Hierarchical neighborhood topology for privacy enhanced collaborative filtering. Proceedings of PEP06, CHI 2006 Workshop on Privacy-Enhanced Personalization, Montreal, Canada pp. 6–13 (2006)
23. Berkovsky, S., Eytani, Y., Kuflik, T., Ricci, F.: Enhancing privacy and preserving accuracy of a distributed collaborative filtering. In: RecSys, pp. 9–16 (2007)
24. Berkovsky, S., Kuflik, T., Ricci, F.: The impact of data obfuscation on the accuracy of collaborative filtering. Expert Syst. Appl. **39**(5), 5033–5042 (2012)
25. Besmer, A., Watson, J., Lipford, H.R.: The impact of social navigation on privacy policy configuration. In: Proceedings of the Sixth Symposium on Usable Privacy and Security, p. Article 7. Redmond, Washington (2010). DOI 10.1145/1837110.1837120
26. Bojars, U., Passant, A., Breslin, J.G., Decker, S.: Social network and data portability using semantic web technologies. In: 2nd Workshop on Social Aspects of the Web (SAW 2008) at BIS2008, pp. 5–19 (2008)

27. Bonneau, J., Anderson, J., Danezis, G.: Prying data out of a social network. In: Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances in, pp. 249–254. IEEE (2009)
28. Brown, C.L., Krishna, A.: The skeptical shopper: A metacognitive account for the effects of default options on choice. Journal of Consumer Research **31**(3), 529–539 (2004). DOI 10.1086/425087
29. Buley, T.: Netflix settles privacy lawsuit, cancels prize sequel. Forbes (3 December 2010).
30. Bustos, L.: Best practice gone bad: 4 shocking A/B tests (2012). URL http://www.getelastic.com/best-practice-gone-bad-4-shocking-ab-tests/
31. Calandrino, J.A., Kilzer, A., Narayanan, A., Felten, E.W., Shmatikov, V.: "you might also like:" privacy risks of collaborative filtering. In: IEEE Symposium on Security and Privacy, pp. 231–246 (2011)
32. Canny, J.F.: Collaborative filtering with privacy. In: Proceedings of the IEEE Symposium on Security and Privacy, pp. 45–57. IEEE Computer Society, Washington, DC, USA (2002)
33. Canny, J.F.: Collaborative filtering with privacy via factor analysis. In: SIGIR, pp. 238–245 (2002)
34. Cavoukian, A.: Privacy by design: The 7 foundational principles. Tech. rep., Information and Privacy Commissioner of Ontario, Canada, Ontario, Canada (2009).
35. Cavusoglu, H., Phan, T., Cavusoglu, H.: Privacy controls and content sharing patterns of online social network users: A natural experiment. ICIS 2013 Proceedings (2013)
36. Chaabane, A., Acs, G., Kaafar, M.A.: You Are What You Like! Information Leakage Through Users' Interests. In: 19th Annual Network & Distributed System Security Symposium (2012)
37. Chellappa, R.K., Sin, R.G.: Personalization versus privacy: An empirical examination of the online consumer's dilemma. Information Technology and Management **6**(2), 181–202 (2005). DOI 10.1007/s10799-005-5879-y
38. Chen, T., Boreli, R., Kaafar, D., Friedman, A.: On the effectiveness of obfuscation techniques in online social networks. In: The 14th Privacy Enhancing Technologies Symposium, pp. 42–62 (2014)
39. Cissée, R., Albayrak, S.: An agent-based approach for privacy-preserving recommender systems. In: Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS '07, pp. 182:1–182:8. ACM, New York, NY, USA (2007)
40. Compañó, R., Lusoli, W.: The policy maker's anguish: Regulating personal data behavior between paradoxes and dilemmas. In: T. Moore, D. Pym, C. Ioannidis (eds.) Economics of Information Security and Privacy, pp. 169–185. Springer US, New York, NY (2010)
41. Consolvo, S., Smith, I., Matthews, T., LaMarca, A., Tabert, J., Powledge, P.: Location disclosure to social relations: why, when, & what people want to share. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 81–90. Portland, OR (2005). DOI 10.1145/1054972.1054985
42. Consumer Reports: Facebook & your privacy: Who sees the data you share on the biggest social network? (2012). URL http://www.consumerreports.org/cro/magazine/2012/06/facebook-your-privacy
43. Cranor, L.F.: Web Privacy with P3P. O'Reilly & Associates, Inc., Sebastopol, CA (2002)
44. Cranor, L.F.: 'I didn't buy it for myself': privacy and ecommerce personalization. In: WPES, pp. 111–117 (2003)
45. Culnan, M.J., Bies, R.J.: Consumer privacy: Balancing economic and justice considerations. Journal of Social Issues **59**(2), 323–342 (2003). DOI 10.1111/1540-4560.00067
46. Dhar, V., Hsieh, J., Sundararajan, A.: Comments on 'Protecting consumer privacy in an era of rapid change: Aproposed framework for businesses and policymakers'. NYU Working Paper CEDER-11-04, New York University, New York, NY (2011)
47. Duhigg, C.: How companies learn your secrets. New York Times Magazine (16 February 2012). URL http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html. [Online; accessed 22-January-2014]
48. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: TCC, pp. 265–284 (2006)

49. Dwork, C., Pitassi, T., Naor, M., Rothblum, G.N.: Differential privacy under continual observation. In: STOC, pp. 715–724 (2010)
50. Egelman, S., Tsai, J., Cranor, L.F., Acquisti, A.: Timing is everything?: the effects of timing and placement of online privacy indicators. In: Proceedings of the 27th international conference on Human factors in computing systems, CHI '09, pp. 319–328. ACM (2009)
51. Erkin, Z., Beye, M., Veugen, T., Lagendijk, R.L.: Privacy enhanced recommender system. Thirty-first Symposium on Information Theory in the Benelux pp. 35–42 (2010)
52. Erkin, Z., Veugen, T., Lagendijk, R.L.: Generating private recommendations in a social trust network. In: CASoN, pp. 82–87 (2011)
53. Erkin, Z., Veugen, T., Toft, T., Lagendijk, R.L.: Generating private recommendations efficiently using homomorphic encryption and data packing. IEEE Transactions on Information Forensics and Security **7**(3), 1053–1066 (2012)
54. EU: Directive 2002/58/EC of the european parliament and of the council concerning the processing of personal data and the protection of privacy in the electronic communications sector. Tech. rep., European Commission (2002)
55. EU: Proposal for a directive of the european parliament and of the council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data. Tech. Rep. 2012/0010 (COD), European Commission (2012)
56. Felten, E.W.: Understanding trusted computing: Will its benefits outweigh its drawbacks? IEEE Security & Privacy **1**(3), 60–62 (2003)
57. Friedman, A., Sharfman, I., Keren, D., Schuster, A.: Privacy-preserving distributed stream monitoring. In: Proceedings of the 21st Annual Network & Distributed System Security Symposium, NDSS '14. Internet Society (2014)
58. van de Garde-Perik, E., Markopoulos, P., de Ruyter, B., Eggen, B., Ijsselsteijn, W.: Investigating privacy attitudes and behavior in relation to personalization. Social Science Computer Review **26**(1), 20–43 (2008). DOI 10.1177/0894439307307682
59. Gardner, J.: 12 surprising A/B test results to stop you making assumptions (2012). URL http://unbounce.com/a-b-testing/shocking-results/
60. Goldfarb, A., Tucker, C.E.: Privacy regulation and online advertising. Management Science **57**(1), 57–71 (2011). DOI 10.1287/mnsc.1100.1246
61. Hancock, J.T., Thom-Santelli, J., Ritchie, T.: Deception and design: the impact of communication technology on lying behavior. In: Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '04, pp. 129–134. ACM, New York, NY, USA (2004)
62. Hann, I.H., Hui, K.L., Lee, S.Y., Png, I.: Overcoming online information privacy concerns: An information-processing theory approach. Journal of Management Information Systems **24**(2), 13–42 (2007). DOI 10.2753/MIS0742-1222240202
63. Harris, L., Westin, A.F., associates: Consumer privacy attitudes: A major shift since 2000 and why. Tech. Rep. 10, Harris Interactive, Inc. (2003)
64. Heitmann, B., Kim, J.G., Passant, A., Hayes, C., Kim, H.G.: An architecture for privacy-enabled user profile portability on the web of data. In: Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems, HetRec '10, pp. 16–23. ACM, New York, NY, USA (2010). DOI 10.1145/1869446.1869449
65. Hoens, T.R., Blanton, M., Chawla, N.V.: A private and reliable recommendation system for social networks. In: Proceedings of the 2010 IEEE Second International Conference on Social Computing, SOCIALCOM '10, pp. 816–825. IEEE Computer Society, Washington, DC, USA (2010). DOI 10.1109/SocialCom.2010.124
66. Hollenbach, J., Presbrey, J., Berners-Lee, T.: Using rdf metadata to enable access control on the social semantic web. In: Proceedings of the Workshop on Collaborative Construction, Management and Linking of Structured Knowledge (CK2009), vol. 514 (2009)
67. House, W.: Consumer data privacy in a networked world: A framework for protecting privacy and promoting innovation in the global economy. Tech. rep., White House, Washington, D.C. (2012)

68. Hui, K.L., Teo, H.H., Lee, S.Y.T.: The value of privacy assurance: An exploratory field experiment. MIS Quarterly **31**(1), 19–33 (2007)
69. ICO: Guidance on the rules on use of cookies and similar technologies. Tech. rep., Information Commissioner's Office (2012)
70. Isaacman, S., Ioannidis, S., Chaintreau, A., Martonosi, M.: Distributed rating prediction in user generated content streams. In: RecSys, pp. 69–76 (2011)
71. Jeckmans, A., Tang, Q., Hartel, P.: Privacy-preserving collaborative filtering based on horizontally partitioned dataset. In: Collaboration Technologies and Systems (CTS), 2012 International Conference on, pp. 439–446 (2012). DOI 10.1109/CTS.2012.6261088
72. Jeckmans, A.J., Beye, M., Erkin, Z., Hartel, P., Lagendijk, R.L., Tang, Q.: Privacy in recommender systems. In: Social Media Retrieval, Computer Communications and Networks, pp. 263–281. Springer (2013)
73. John, L.K., Acquisti, A., Loewenstein, G.: Strangers on a plane: Context-dependent willingness to divulge sensitive information. Journal of consumer research **37**(5), 858–873 (2011)
74. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decis. Support Syst. **43**(2), 618–644 (2007). DOI 10.1016/j.dss.2005.05.019
75. Kahneman, D., Tversky, A.: Prospect theory: An analysis of decision under risk. Econometrica **47**(2), 263–292 (1979). DOI 10.2307/1914185
76. Kaleli, C., Polat, H.: Providing private recommendations using naïve bayesian classifier. In: K.M. Wegrzyn-Wolska, P.S. Szczepaniak (eds.) Advances in Intelligent Web Mastering, *Advances in Soft Computing*, vol. 43, pp. 168–173. Springer Berlin Heidelberg (2007)
77. Kandappu, T., Friedman, A., Boreli, R., Sivaraman, V.: PrivacyCanary: Privacy-aware recommenders with adaptive input obfuscation. In: MASCOTS (2014)
78. Knijnenburg, B.P., Jin, H.: The persuasive effect of privacy recommendations. In: Twelfth Annual Workshop on HCI Research in MIS, p. Paper 16. Milan, Italy (2013)
79. Knijnenburg, B.P., Kobsa, A.: Helping users with information disclosure decisions: potential for adaptation. In: Proceedings of the 2013 ACM international conference on Intelligent User Interfaces, pp. 407–416. ACM Press, Santa Monica, CA (2013)
80. Knijnenburg, B.P., Kobsa, A.: Making decisions about privacy: Information disclosure in context-aware recommender systems. ACM Transactions on Interactive Intelligent Systems **3**(3), 20:1–20:23 (2013). DOI 10.1145/2499670
81. Knijnenburg, B.P., Kobsa, A., Jin, H.: Counteracting the negative effect of form auto-completion on the privacy calculus. In: ICIS 2013 Proceedings. Milan, Italy (2013)
82. Knijnenburg, B.P., Kobsa, A., Jin, H.: Dimensionality of information disclosure behavior. International Journal of Human-Computer Studies **71**(12), 1144–1162 (2013)
83. Knijnenburg, B.P., Kobsa, A., Jin, H.: Preference-based location sharing: are more privacy options really better? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2667–2676. ACM, Paris, France (2013). DOI 10.1145/2470654.2481369
84. Knijnenburg, B.P., Kobsa, A., Saldamli, G.: Privacy in mobile personalized systems: The effect of disclosure justifications. In: Proceedings of the SOUPS 2012 Workshop on Usable Privacy & Security for Mobile Devices, pp. 11:1–11:5. Washington, DC (2012)
85. Knijnenburg, B.P., Willemsen, M.C., Hirtbach, S.: Receiving recommendations and providing feedback: The user-experience of a recommender system. In: EC-Web, pp. 207–216 (2010)
86. Kobsa, A.: Tailoring privacy to users' needs (invited keynote). In: M. Bauer, P.J. Gmytrasiewicz, J. Vassileva (eds.) User Modeling 2001, no. 2109 in Lecture Notes in Computer Science, pp. 303–313. Springer Verlag (2001).
87. Kobsa, A.: Privacy-enhanced web personalization. In: P. Brusilovsky, A. Kobsa, W. Nejdl (eds.) The Adaptive Web, pp. 628–670. Springer-Verlag, Berlin, Heidelberg (2007)
88. Kobsa, A., Knijnenburg, B.P., Livshits, B.: Let's do it at my place instead? attitudinal and behavioral study of privacy in client-side personalization. In: ACM CHI Conference on Human Factors in Computing Systems. Toronto, Canada (2014)
89. Kobsa, A., Schreck, J.: Privacy through pseudonymity in user-adaptive systems. ACM Trans. Internet Techn. **3**(2), 149–183 (2003)

90. Kobsa, A., Teltzrow, M.: Contextualized communication of privacy practices and personalization benefits: Impacts on users' data sharing and purchase behavior. In: D. Martin, A. Serjantov (eds.) Privacy Enhancing Technologies: Revised Selected Papers of the 4th International Workshop, PET 2004, Toronto, Canada, May 26–28, 2004, *LNCS*, vol. 3424, pp. 329–343. Springer Berlin Heidelberg (2005). DOI 10.1007/b136164
91. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences (2013)
92. Krishnamurthy, B., Wills, C.: Privacy diffusion on the web: A longitudinal perspective. In: Proceedings of the 18th International Conference on World Wide Web, WWW '09, pp. 541–550. ACM, New York, NY, USA (2009). DOI 10.1145/1526709.1526782
93. Lai, Y.L., Hui, K.L.: Internet opt-in and opt-out: Investigating the roles of frames, defaults and privacy concerns. In: Proceedings of the 2006 ACM SIGMIS CPR Conference on Computer Personnel Research, pp. 253–263. Claremont, CA (2006). DOI 10.1145/1125170.1125230
94. Lathia, N., Hailes, S., Capra, L.: Private distributed collaborative filtering using estimated concordance measures. In: Proceedings of the 2007 ACM Conference on Recommender Systems, RecSys '07, pp. 1–8. ACM, New York, NY, USA (2007)
95. Laufer, R.S., Proshansky, H.M., Wolfe, M.: Some analytic dimensions of privacy. In: R. Küller (ed.) Proceedings of the Lund Conference on Architectural Psychology. Dowden, Hutchinson & Ross, Lund, Sweden (1973)
96. Lemire, D., Maclachlan, A.: Slope one predictors for online rating-based collaborative filtering. In: SDM (2005)
97. Li, H., Sarathy, R., Xu, H.: Understanding situational online information disclosure as a privacy calculus. Journal of Computer Information Systems **51**(1), 62–71 (2010)
98. Li, T., Unger, T.: Willing to pay for quality personalization? trade-off between quality and privacy. European Journal of Information Systems **21**(6), 621–642 (2012)
99. Lin, J., Sadeh, N.M., Amini, S., Lindqvist, J., Hong, J.I., Zhang, J.: Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In: UbiComp, pp. 501–510 (2012)
100. Liu, Y., Gummadi, K.P., Krishnamurthy, B., Mislove, A.: Analyzing facebook privacy settings: user expectations vs. reality. In: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, pp. 61–70. ACM, Berlin, Germany (2011)
101. Machanavajjhala, A., Korolova, A., Sarma, A.D.: Personalized social recommendations - accurate or private? PVLDB **4**(7), 440–450 (2011)
102. Malhotra, N.K., Kim, S.S., Agarwal, J.: Internet users' information privacy concerns (IUIPC): the construct, the scale, and a nomological framework. Information Systems Research **15**(4), 336–355 (2004). DOI 10.1287/isre.1040.0032
103. McSherry, F., Mironov, I.: Differentially private recommender systems: Building privacy into the netflix prize contenders. In: KDD, pp. 627–636 (2009)
104. Metzger, M.J.: Privacy, trust, and disclosure: Exploring barriers to electronic commerce. Journal of Computer-Mediated Communication **9**(4) (2004)
105. Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N.: Detecting price and search discrimination on the internet. In: HotNets, pp. 79–84 (2012)
106. Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N.: Crowd-assisted search for price discrimination in e-commerce: first results. In: CoNEXT, pp. 1–6 (2013)
107. Miller, B.N., Konstan, J.A., Riedl, J.: Pocketlens: Toward a personal recommender system. ACM Transactions on Information Systems **22**(3), 437–476 (2004)
108. Narayanan, A., Shmatikov, V.: Robust de-anonymization of large sparse datasets. In: IEEE Symposium on Security and Privacy, pp. 111–125 (2008)
109. Nikolaenko, V., Ioannidis, S., Weinsberg, U., Joye, M., Taft, N., Boneh, D.: Privacy-preserving matrix factorization. In: ACM Conference on Computer and Communications Security, pp. 801–812 (2013)
110. Nissenbaum, H.: Privacy as contextual integrity. Washington Law Review **79**(1), 101–139 (2004)
111. Nissenbaum, H.: A contextual approach to privacy online. Daedalus **140**(4), 32–48 (2011)

112. Nissim, K., Raskhodnikova, S., Smith, A.: Smooth sensitivity and sampling in private data analysis. In: Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing, STOC '07, pp. 75–84. ACM, New York, NY, USA (2007)
113. Norberg, P.A., Horne, D.R., Horne, D.A.: The privacy paradox: Personal information disclosure intentions versus behaviors. Journal of Consumer Affairs **41**(1), 100–126 (2007)
114. OECD: Recommendation of the council concerning guidelines governing the protection of privacy and transborder flows of personal data. Tech. rep., Organization for Economic Co-operation and Development (1980). Print file://Lit1/OECD-privacy-1980.htm
115. Ohm, P.: Broken promises of privacy: Responding to the surprising failure of anonymization. UCLA Law Review **57**, 1701 (2010)
116. Paillier, P.: Public-key cryptosystems based on composite degree residuosity classes. In: EUROCRYPT, pp. 223–238 (1999)
117. Pallapa, G., Das, S.K., Di Francesco, M., Aura, T.: Adaptive and context-aware privacy preservation exploiting user interactions in smart environments. Pervasive and Mobile Computing **12**, 232–243 (2014). DOI 10.1016/j.pmcj.2013.12.004. URL http://www.sciencedirect.com/science/article/pii/S1574119213001557
118. Parameswaran, R., Blough, D.M.: Privacy preserving collaborative filtering using data obfuscation. In: Proceedings of the IEEE Conference on Granular Computing, p. 380 (2007)
119. Pfitzmann, A., Hansen, M.: A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf (2010)
120. Polat, H., Du, W.: Privacy-preserving collaborative filtering using randomized perturbation techniques. In: ICDM, pp. 625–628 (2003)
121. Polat, H., Du, W.: Svd-based collaborative filtering with privacy. In: SAC, pp. 791–795 (2005)
122. Polat, H., Du, W.: Achieving private recommendations using randomized response techniques. In: PAKDD, pp. 637–646 (2006)
123. Put, A., Dacosta, I., Milutinovic, M., De Decker, B., Seys, S., Boukayoua, F., Naessens, V., Vanhecke, K., De Pessemier, T., Martens, L.: inshopnito: An advanced yet privacy-friendly mobile shopping application. In: Proceedings of 2014 IEEE World Congress on Services. IEEE Computer Society Press (2014). URL https://lirias.kuleuven.be/handle/123456789/454582
124. Ramakrishnan, N., Keller, B.J., Mirza, B.J., Grama, A., Karypis, G.: Privacy risks in recommender systems. IEEE Internet Computing **5**(6), 54–62 (2001)
125. Ravichandran, R., Benisch, M., Kelley, P., Sadeh, N.: Capturing social networking privacy preferences:. In: I. Goldberg, M. Atallah (eds.) Privacy Enhancing Technologies, *Lecture Notes in Computer Science*, vol. 5672, pp. 1–18. Springer Berlin / Heidelberg (2009).
126. Renckes, S., Polat, H., Oysal, Y.: A new hybrid recommendation algorithm with privacy. Expert Systems **29**(1), 39–55 (2012)
127. Riboni, D., Bettini, C.: Private context-aware recommendation of points of interest: An initial investigation. In: PerCom Workshops, pp. 584–589 (2012)
128. Rifon, N.J., LaRose, R., Choi, S.M.: Your privacy is sealed: Effects of web privacy seals on trust and personal disclosures. Journal of Consumer Affairs **39**(2), 339–360 (2005)
129. Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M., Rao, J.: Understanding and capturing people's privacy policies in a mobile social networking application. Personal and Ubiquitous Computing **13**(6), 401–412 (2009). DOI 10.1007/s00779-008-0214-3
130. Schafer, J.B., Konstan, J.A., Riedl, J.: E-commerce recommendation applications. Data Min. Knowl. Discov. **5**(1/2), 115–153 (2001)
131. Shokri, R., Pedarsani, P., Theodorakopoulos, G., Hubaux, J.P.: Preserving privacy in collaborative filtering through distributed aggregation of offline profiles. In: RecSys, pp. 157–164 (2009)
132. Smith, H.J., Dinev, T., Xu, H.: Information privacy research: An interdisciplinary review. MIS Quarterly **35**(4), 989–1016 (2011)
133. Smith, H.J., Milberg, S.J., Burke, S.J.: Information privacy: Measuring individuals' concerns about organizational practices. MIS Quarterly **20**(2), 167–196 (1996)

134. Smith, N.C., Goldstein, D.G., Johnson, E.J.: Choice without awareness: Ethical and policy implications of defaults. Journal of Public Policy & Marketing **32**(2), 159–172 (2013)
135. Solove, D.J.: Privacy self-management and the consent dilemma. Harvard Law Review **126**, 1880–1903 (2013)
136. Stafford, J., Wallnau, K.: Is third party certification necessary. In: Proceedings of the 4th ICSE Workshop on Component-based Software Engineering: Component Certification and System Prediction, pp. 13–17 (2001)
137. Suguri, H.: A standardization effort for agent technologies: The foundation for intelligent physical agents and its activities. In: HICSS (1999)
138. Taylor, D., Davis, D., Jillapalli, R.: Privacy concern and online personalization: The moderating effects of information control and compensation. Electronic Commerce Research **9**(3), 203–223 (2009). DOI 10.1007/s10660-009-9036-2
139. Toch, E., Wang, Y., Cranor, L.F.: Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. User Modeling and User-Adapted Interaction **22**(1–2), 203–220 (2012). DOI 10.1007/s11257-011-9110-z
140. TRUSTe: First in-depth analysis of the impact of EU cookie directive shows majority of users choosing to allow advertising cookies (2012).
141. Vallet, D., Friedman, A., Berkovsky, S.: Matrix factorization without user data retention. In: PAKDD (2014)
142. Walton, D.: Plausible deniability and evasion of burden of proof. Argumentation **10**(1), 47–58 (1996). DOI 10.1007/BF00126158
143. Wang, W., Benbasat, I.: Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. Journal of Management Information Systems **23**(4), 217–246 (2007). DOI 10.2753/MIS0742-1222230410
144. Wang, Y., Kobsa, A.: Impacts of privacy laws and regulations on personalized systems. In: A. Kobsa, R. Chellappa, S. Spiekermann (eds.) Proceedings of PEP06, CHI 2006 Workshop on Privacy-Enhanced Personalization, pp. 44–46. Springer Verlag, Montréal, Canada (2006)
145. Wang, Y., Kobsa, A.: Respecting users' individual privacy constraints in web personalization. In: C. Conati, K. McCoy, G. Paliouras (eds.) User Modeling 2007, pp. 157–166. Springer Verlag (2007)
146. Wang, Y., Leon, P.G., Scott, K., Chen, X., Acquisti, A., Cranor, L.F.: Privacy nudges for social media: An exploratory facebook study. In: Second International Workshop on Privacy and Security in Online Social Media. Rio De Janeiro, Brazil (2013)
147. Weinsberg, U., Bhagat, S., Ioannidis, S., Taft, N.: Blurme: inferring and obfuscating user gender based on ratings. In: RecSys, pp. 195–202 (2012)
148. Westin, A.F., Harris, L., associates: The Dimensions of privacy : a national opinion research survey of attitudes toward privacy. Garland Publishing, New York (1981)
149. Xu, H., Luo, X.R., Carroll, J.M., Rosson, M.B.: The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing. Decision Support Systems **51**(1), 42–52 (2011). DOI 10.1016/j.dss.2010.11.017
150. Xu, H., Teo, H.H., Tan, B.C.Y.: Predicting the adoption of location-based services: The role of trust and perceived privacy risk. In: Proceedings of the International Conference on Information Systems, pp. 861–874. Las Vegas, NV (2005)
151. Xu, H., Teo, H.H., Tan, B.C.Y., Agarwal, R.: Effects of individual self-protection, industry self-regulation, and government regulation on privacy concerns: A study of location-based services. Information Systems Research (2012). DOI 10.1287/isre.1120.0416
152. Xu, H., Wang, N., Grossklags, J.: Privacy-by-ReDesign: alleviating privacy concerns for third-party applications. In: ICIS 2012 Proceedings. Orlando, FL (2012)
153. Yakut, I., Polat, H.: Privacy-preserving eigentaste-based collaborative filtering. In: A. Miyaji, H. Kikuchi, K. Rannenberg (eds.) Advances in Information and Computer Security, *Lecture Notes in Computer Science*, vol. 4752, pp. 169–184. Springer Berlin Heidelberg (2007)
154. Yakut, I., Polat, H.: Arbitrarily distributed data-based recommendations with privacy. Data & Knowledge Engineering **72**(0), 239–256 (2012)

155. Yao, A.C.: Protocols for secure computations. In: Proceedings of the 23rd Annual Symposium on Foundations of Computer Science, SFCS '82, pp. 160–164. IEEE Computer Society, Washington, DC, USA (1982). DOI 10.1109/SFCS.1982.88
156. Zhang, A., Bhamidipati, S., Fawaz, N., Kveton, B.: Priview: Media consumption and recommendation meet privacy against inference attacks. In: W2SP (2014)
157. Zheleva, E., Getoor, L.: To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In: WWW, pp. 531–540 (2009)
158. Zhu, T., Ren, Y., Zhou, W., Rong, J., Xiong, P.: An effective privacy preserving algorithm for neighborhood-based collaborative filtering. Future Generation Computer Systems (2013)