Privacy and Personalization

Claudia Lopez

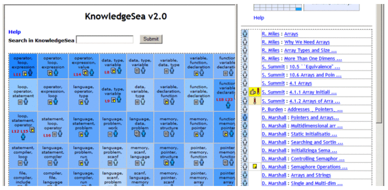
Is someone concerned about privacy when learning about personalized systems?

- "This is the reason that Google has been criticized by many people. They consider user profiling as an activity which may violate their privacy." (Jia Wu)
- "What I concerned is the privacy problem. What if I ever shared one webpage but I do not want it to be presented to other users?" (Jiangyue Zhu)
- "…privacy is often one of the problems that comes into socially powered things from searches to networking...too much exposure effects the general sense of safety, causing individuals to not want to share…" (Melissa Dukes)
- 18 comments related to privacy in our discussion form

What I'll be telling you the next 48 minutes?

- What are the privacy behaviors and risks that users perceive in (web) personalized systems?
- What factors influence users' willingness to disclose personal information?
- How laws and regulation protect users' privacy?
- Mechanisms to build personalized systems that protect users' privacy
- Hope we all learn about privacy from both perspectives, as users and as developers

So, what is the problem?



KnowledgeSea v2.0

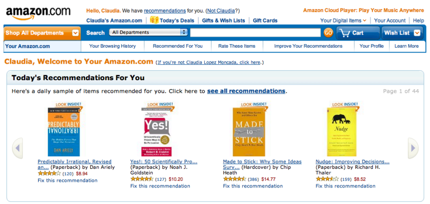
Help

Search in KnowledgeSea Submit

operator, loop, expression L11	operator, loop, expression	operator, expression, value L14	data, type, variable L8	data, type, variable	variable, data, type	variable, function, declaration	function variable declarati
loop, operator, statement	operator, expression, loop	language, operator, type	data, type, variable L9	data, variable, type	variable, declaration, function	function, variable, declaration	function variable declarati L18 L23
loop, statement, operator L12 L15	statement, loop, operator L16	language, statement, problem	language, problem, work	language, data, problem	memory, variable, structure	memory, function, pointer	function memory pointer
statement, compiler, loop	language, statement, compiler	language, problem, run	language, problem, scanf	memory, scanf, language	memory, pointer, structure	pointer, memory, function	pointer memory functio
file, compiler, include	compiler, file, language	language, compiler, run	language, scanf, problem	scanf, language, memory	memory, pointer, scanf	pointer, memory, array	pointer memory array

Help

R. Miles : Arrays
R. Miles : Why We Need Arrays
R. Miles : Array Types and Size ...
R. Miles : More Than One Dimens ...
S. Summit : 10.5 ``Equivalence'' ...
S. Summit : 10.6 Arrays and Poin ...
S. Summit : 4.1 Arrays
S. Summit : 4.1.1 Array Initiali ...
S. Summit : 4.1.2 Arrays of Arra ...
P. Burden : Addresses , Pointers, ...
D. Marshall : Pointers and Arrays...
D. Marshall : Multidimensional arr ...
D. Marshall : Static Initialisatio ...
D. Marshall : Searching and Sortin ...
D. Marshall : Initializinga Sema ...
D. Marshall : Controlling Semaphor ...
D. Marshall : Semaphore Operations ...
D. Marshall : Arrays and Strings
D. Marshall : Single and Multi-dim ...



Hello, Claudia. We have recommendations for you. (Not Claudia?)

Amazon Cloud Player: Play Your Music Anywhere

Claudia's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards

Your Digital Items | Your Account | Help

Shop All Departments | Search All Departments | GO | Cart | Wish List

Your Amazon.com | Your Browsing History | Recommended For You | Rate These Items | Improve Your Recommendations | Your Profile | Learn More

Claudia, Welcome to Your Amazon.com (If you're not Claudia Lopez Moncada, click here.)

Today's Recommendations For You

Here's a daily sample of items recommended for you. Click here to see all recommendations.

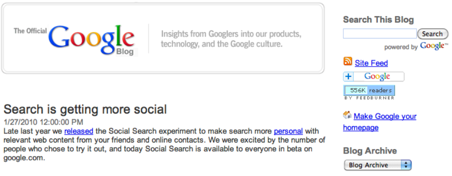
Page 1 of 44

Predictably Irrational, Revised an... (Paperback) by Dan Ariely (120) $8.94
Fix this recommendation

Yes!: 50 Scientifically Pro... (Paperback) by Noah J. Goldstein (127) $10.20
Fix this recommendation

Made to Stick: Why Some Ideas Surv... (Hardcover) by Chip Heath (386) $14.77
Fix this recommendation

Nudge: Improving Decisions... (Paperback) by Richard H. Thaler (159) $8.52
Fix this recommendation



Insights from Googlers into our products, technology, and the Google culture.

Search This Blog

Search

powered by Google™

Site Feed

556K readers BY FEEDBURNER

Make Google your homepage

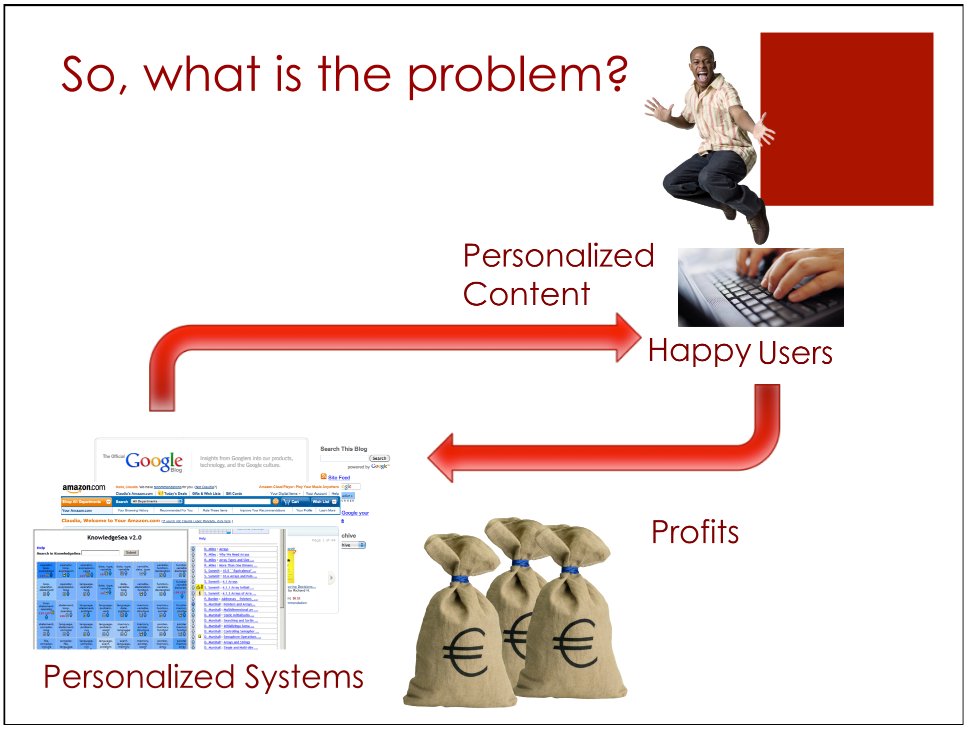
Blog Archive

Blog Archive

Search is getting more social

1/27/2010 12:00:00 PM

Late last year we released the Social Search experiment to make search more personal with relevant web content from your friends and online contacts. We were excited by the number of people who chose to try it out, and today Social Search is available to everyone in beta on google.com.



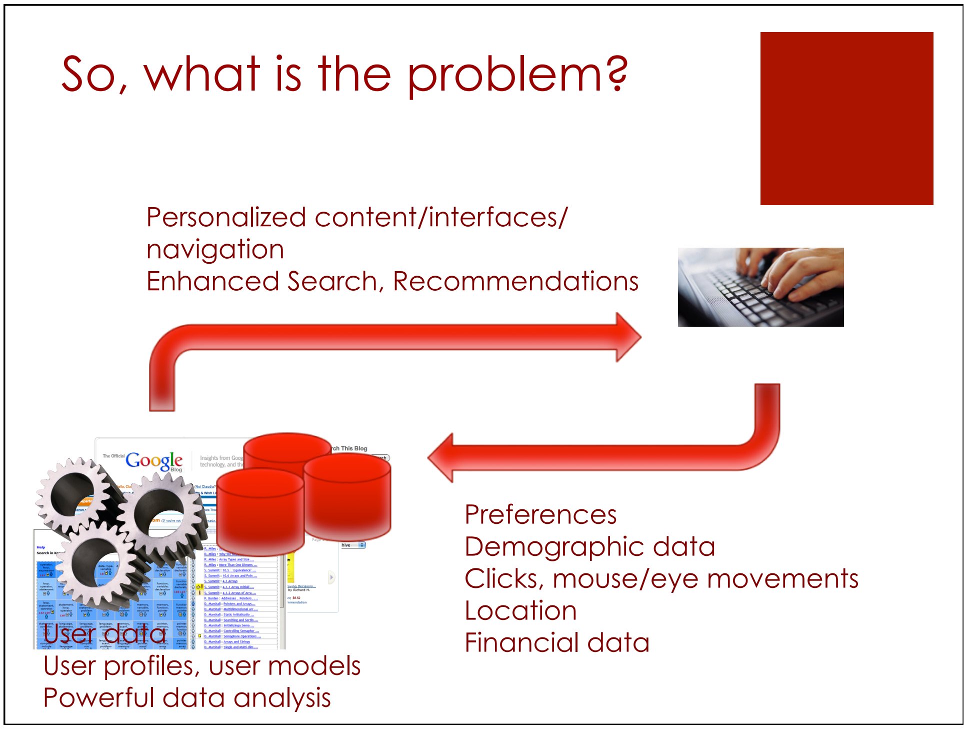
So, what is the problem?

Personalized Content

Happy Users

Profits

Personalized Systems



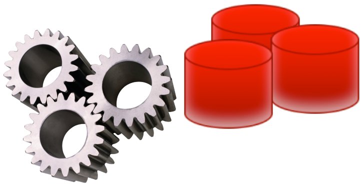
So, what is the problem?

Personalized content/interfaces/ navigation
Enhanced Search, Recommendations

Preferences
Demographic data
Clicks, mouse/eye movements
Location
Financial data

User data
User profiles, user models
Powerful data analysis

Some questions for users



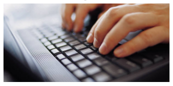
User data
User profiles, user models
Powerful data analysis

- Are the applications protecting our private data?
- Are the applications sharing my information with third-parties or using my data with other purpose?
- Is my private data safe when it is transmitted through the internet?
- Am I willing to disclose some data to enjoy some personalization?
- What are our rights regarding privacy?

Some questions for users

- Can the data about my knowledge in a educational system be used to decide if I would be fired?
- Is Amazon sharing my preferences with third parties? What about my mailing address? My financial info?
- Is Google using the content of my emails for other purposes?
- …..

Some questions for developers



- Are we able to provide personalization and protecting users' privacy?
- How are we going to persuade people to disclose some of their private information?
- Are we respecting the privacy laws?

How can we learn about individual privacy concerns?

- What people says and what people does might be different
- 2 empirical methods to figure it out
  - Inquiry-based methods
    - Ask about privacy attitudes, past behavior, and their anticipated behavior under certain conditions
  - Observation-based methods
    - Observation during empirical studies (e.g. disclosure of data while purchasing products)

Empirical methods are subject to potential biases

- Both methods
  - Biased self-selection
  - Socially desirable responses
  - Discrepancies between stated attitudes and observed behavior
- Inquiry-based methods
  - Answers may not represent reality
- Observation-based methods
  - Results may be not represent behavioral patterns

Findings related to privacy concerns

- Different surveys mostly in US between 1998 and 2003
  - Internet users who
    - are concerned about privacy of info online (70-89.5%)
    - have refused to give personal data ever (82-95%)
    - would never provide personal data (27%)
    - have provided false information (24-40%)
    - are concerned if a business share their data (89-90%)
    - think that sharing info with other site violates privacy (83%)

Findings related to tracking and cookies

- Internet users who
  - are concerned about being tracked (54-63%)
  - think that someone might know the sites they visited(31%
  - feel uncomfortable being tracked across != sites (91%)
  - generally accept cookies (62%)
  - configure their computers to reject cookies (10-25%)
  - delete cookies periodically (53%)

How these findings are related to personalization?

People who decide to lie

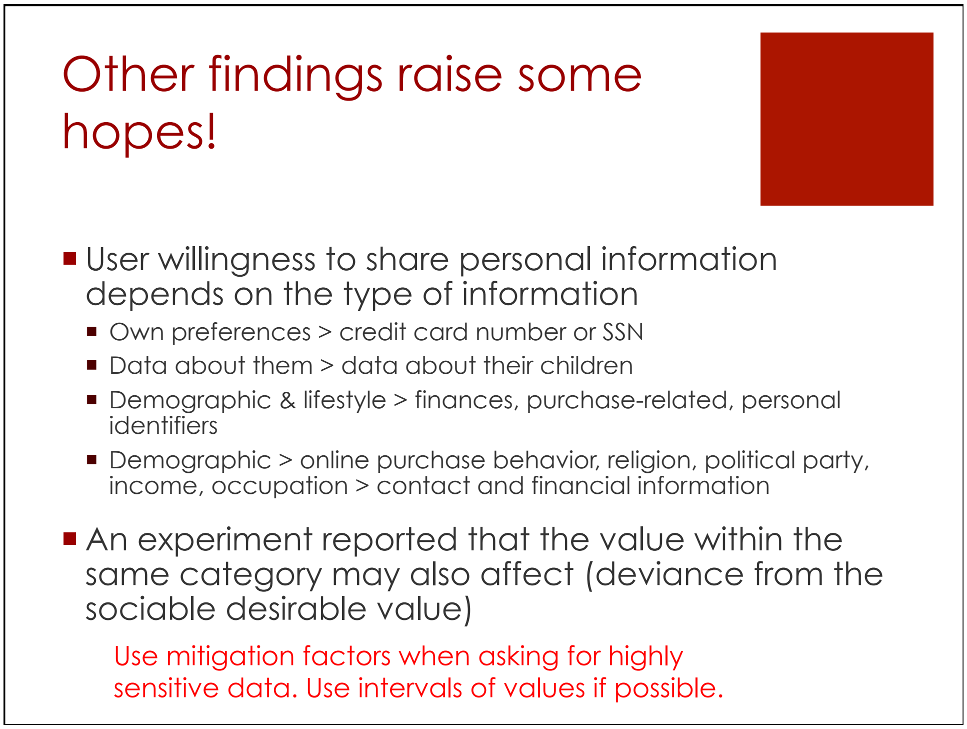
People who doesn't want their data to be shared with other sites, and doesn't trust that sites are not sharing their data

People who doesn't like to be tracked

People who reject or delete cookies

People who won't disclose information

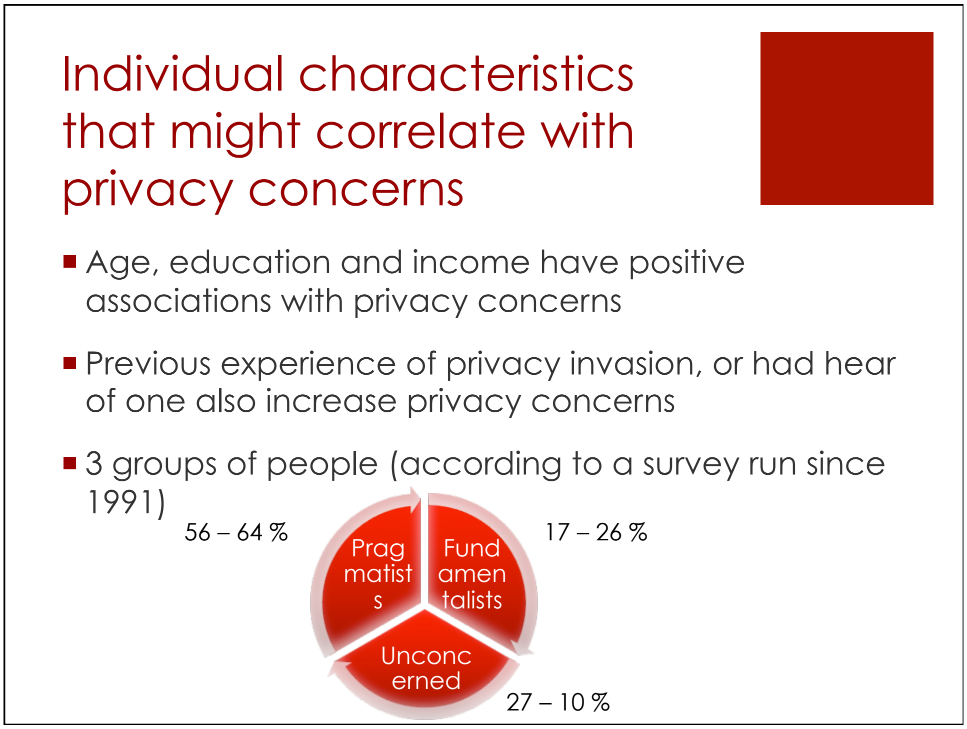
Low level of trust => less disclosure of data
False data => poor user profiles
How to link sessions from the same user (without cookies and registration)?



Other findings raise some hopes!

- User willingness to share personal information depends on the type of information
  - Own preferences > credit card number or SSN
  - Data about them > data about their children
  - Demographic & lifestyle > finances, purchase-related, personal identifiers
  - Demographic > online purchase behavior, religion, political party, income, occupation > contact and financial information
- An experiment reported that the value within the same category may also affect (deviance from the sociable desirable value)

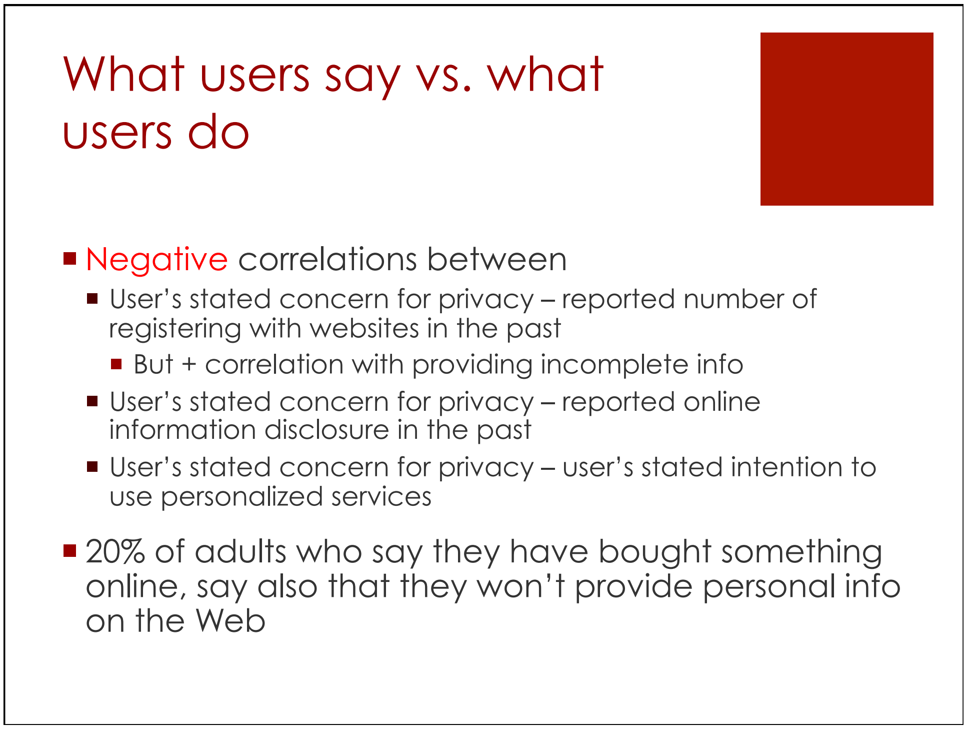
Use mitigation factors when asking for highly sensitive data. Use intervals of values if possible.



Individual characteristics that might correlate with privacy concerns

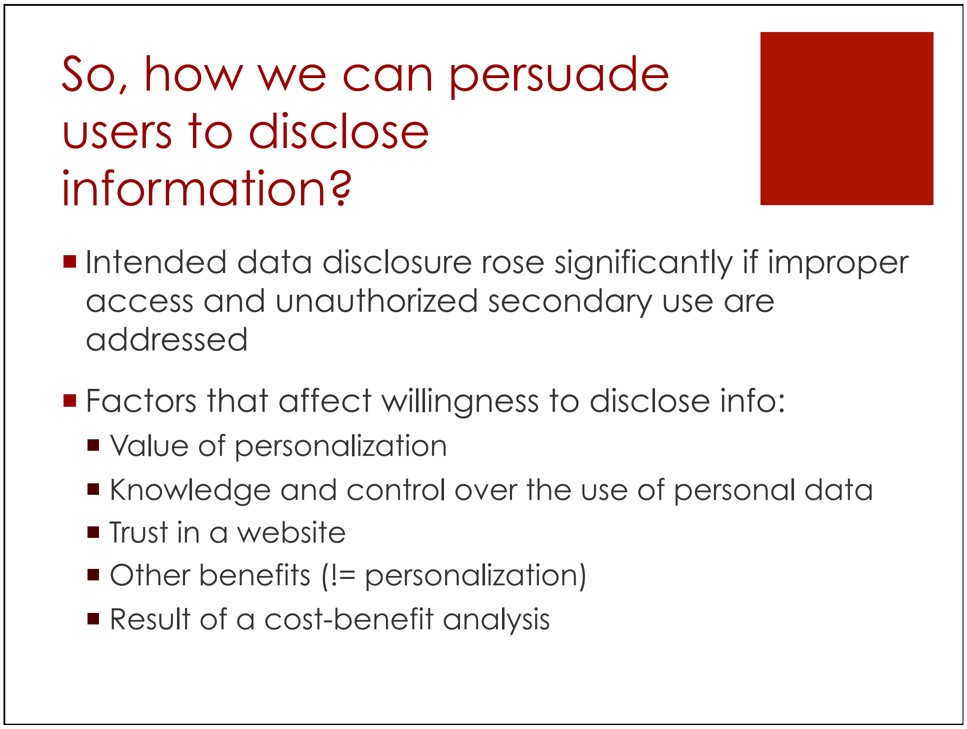
- Age, education and income have positive associations with privacy concerns
- Previous experience of privacy invasion, or had hear of one also increase privacy concerns
- 3 groups of people (according to a survey run since 1991)

56 – 64 %
Prag matist s
Fund amen talists
17 – 26 %
Unconc erned
27 – 10 %



What users say vs. what users do

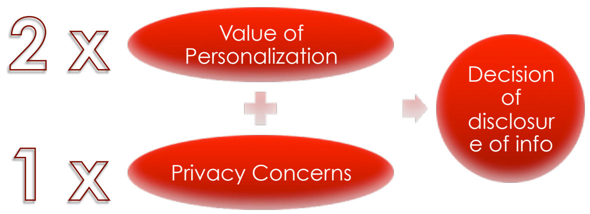
- Negative correlations between
  - User's stated concern for privacy – reported number of registering with websites in the past
    - But + correlation with providing incomplete info
  - User's stated concern for privacy – reported online information disclosure in the past
  - User's stated concern for privacy – user's stated intention to use personalized services
- 20% of adults who say they have bought something online, say also that they won't provide personal info on the Web



So, how we can persuade users to disclose information?

- Intended data disclosure rose significantly if improper access and unauthorized secondary use are addressed
- Factors that affect willingness to disclose info:
  - Value of personalization
  - Knowledge and control over the use of personal data
  - Trust in a website
  - Other benefits (!= personalization)
  - Result of a cost-benefit analysis

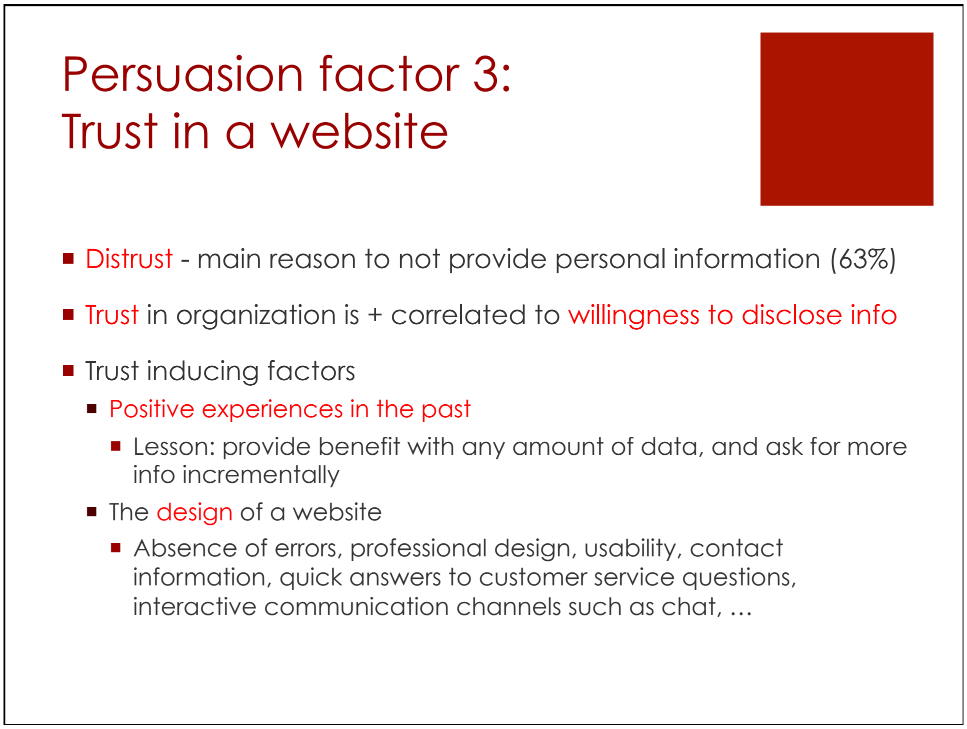
Persuasion factor 1: Value of information



- Does people value personalization?
  - Personalization is a good thing? (yes:59%, no:37%)
  - Willing to give info for personalization (yes: 51&43%, no: 15&39%)
  - It's useful if the site remember basic info (73%), preferences (50%). It's bothering to provide info twice (62%)

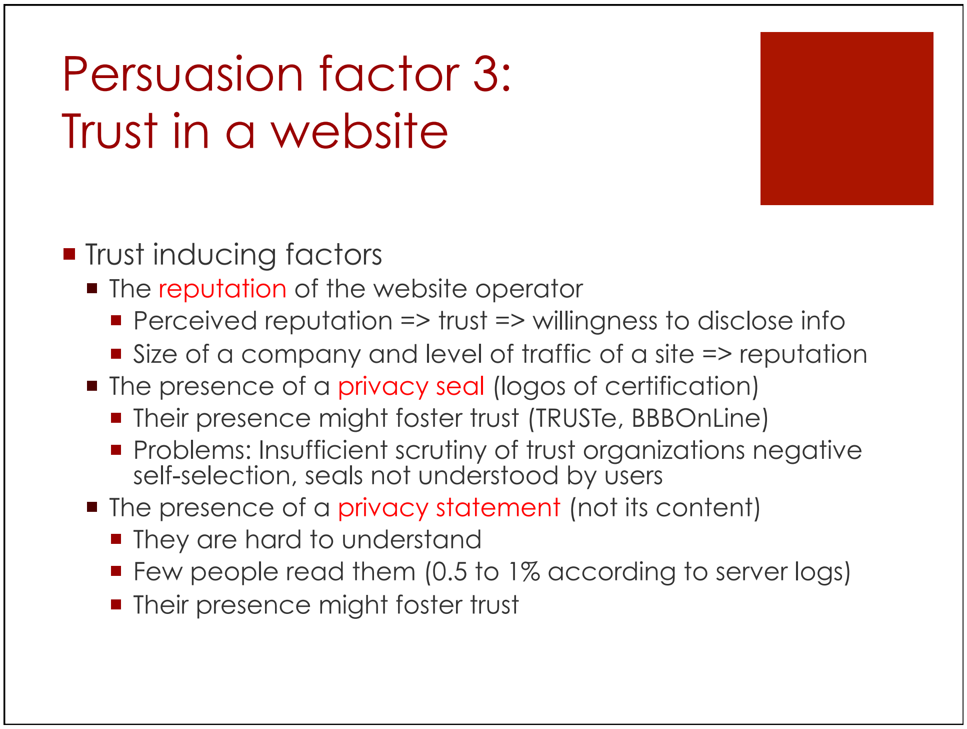
Persuasion factor 2: Knowledge and control over the use of data

- People want to know what personal information will be used (94%, it should be a right 94%) = > trust
  - Failing on it may encourage users to provide false info
- People want to control what personal information is collected (94%) = > trust
  - It's the most important factor in privacy concerns
- An experiment showed that users disclosed more info when they site explained the benefits and privacy practices
- Another experiment allowed user to configure profile settings and adaptation, but it was challenging for users



Persuasion factor 3: Trust in a website

- Distrust - main reason to not provide personal information (63%)
- Trust in organization is + correlated to willingness to disclose info
- Trust inducing factors
  - Positive experiences in the past
    - Lesson: provide benefit with any amount of data, and ask for more info incrementally
  - The design of a website
    - Absence of errors, professional design, usability, contact information, quick answers to customer service questions, interactive communication channels such as chat, …



Persuasion factor 3: Trust in a website

- Trust inducing factors
  - The reputation of the website operator
    - Perceived reputation => trust => willingness to disclose info
    - Size of a company and level of traffic of a site => reputation
  - The presence of a privacy seal (logos of certification)
    - Their presence might foster trust (TRUSTe, BBBOnLine)
    - Problems: Insufficient scrutiny of trust organizations negative self-selection, seals not understood by users
  - The presence of a privacy statement (not its content)
    - They are hard to understand
    - Few people read them (0.5 to 1% according to server logs)
    - Their presence might foster trust

Persuasion factor 4: Benefits other than Personalization

- Financial rewards for personal data disclosure
  - 16% of respondents (US$10) other survey (US$15 - 50)
- Social adjustments benefits
  - The chance to integrate into desired social groups was a significant factor in the decision to provide data
    - face-to-face and online groups were a factor for extrovert people, and only online groups made a difference for introvert people

Persuasion factor 5: Result of cost-benefit analysis

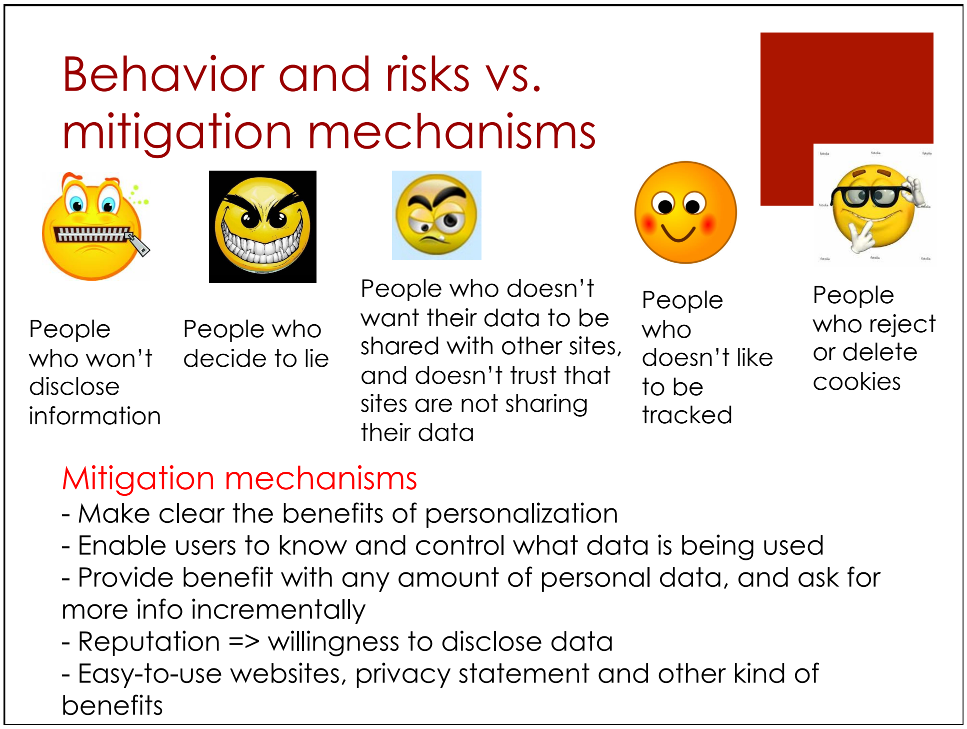
- Cost and benefit are also mediated by other factors such as trust
- However, users may lack information to make educated privacy-related decisions
  - Few people read privacy policies
  - Some overestimate immediate benefits, and underestimate future negative impacts

Lessons to remind

- Make clear the benefits of personalization
- Inform what personal data will be used and allow users to control the data that is being collected (if possible)
- Provide benefit with any amount of data, and ask for more personal info incrementally
- Reputation => willingness to disclose data
- Provide easy-to-use websites + contact info or mech.
- Privacy statement are good, but it might not affect disclosure
- Financial and social benefits might persuade people to provide data

Let’s summarize what we have seen, and what else is coming

- What are the privacy behavior and risks that users perceive in (web) personalized systems?
- What factors influence users’ willingness to disclose personal information?
- How laws and regulation protect users’ privacy?
- Mechanisms to build personalized systems that protect users’ privacy
- Hope we all learn about privacy from both perspectives, as users and as developers



Behavior and risks vs. mitigation mechanisms

People who won't disclose information

People who decide to lie

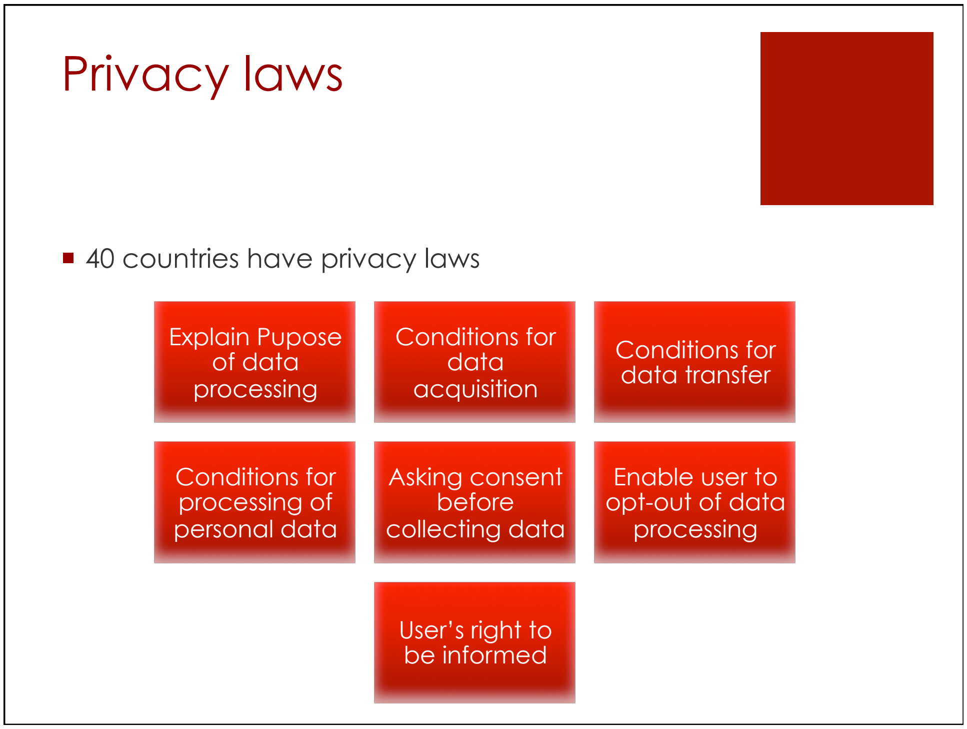
People who doesn't want their data to be shared with other sites, and doesn't trust that sites are not sharing their data

People who doesn't like to be tracked

People who reject or delete cookies

Mitigation mechanisms

- Make clear the benefits of personalization
- Enable users to know and control what data is being used
- Provide benefit with any amount of personal data, and ask for more info incrementally
- Reputation => willingness to disclose data
- Easy-to-use websites, privacy statement and other kind of benefits



Privacy laws

- 40 countries have privacy laws

Explain Pupose of data processing

Conditions for data acquisition

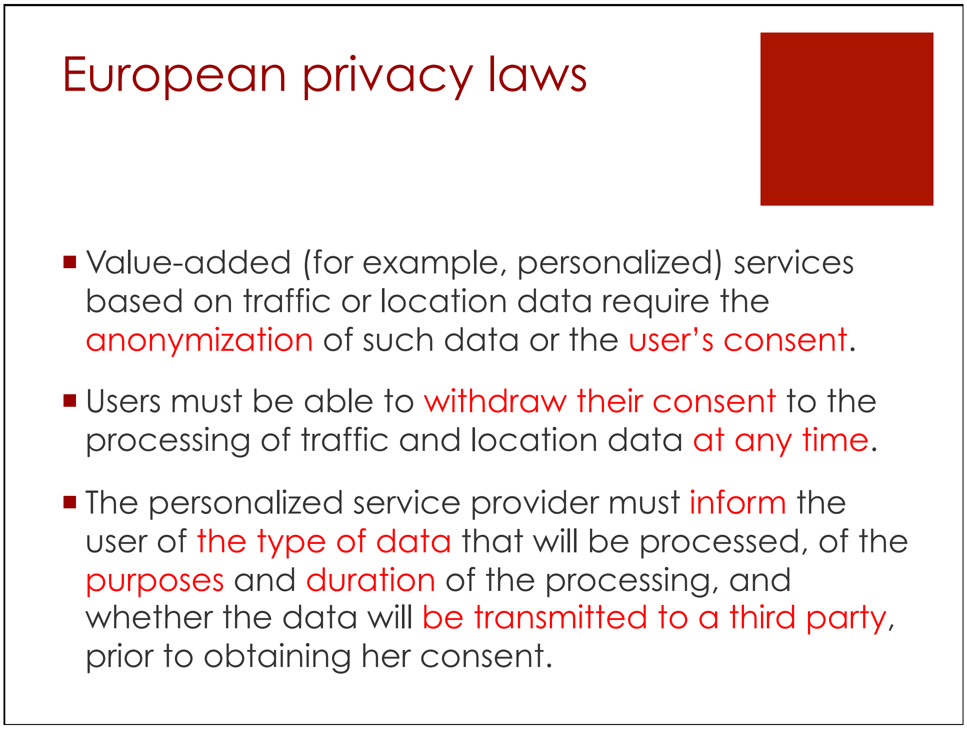
Conditions for data transfer

Conditions for processing of personal data

Asking consent before collecting data

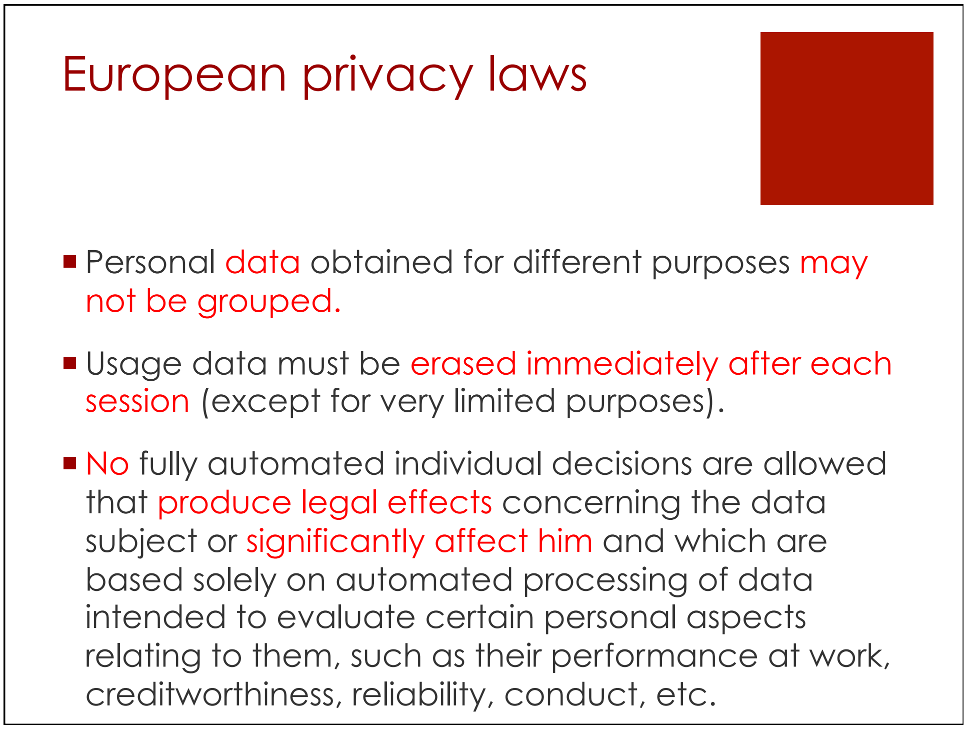
Enable user to opt-out of data processing

User's right to be informed



European privacy laws

- Value-added (for example, personalized) services based on traffic or location data require the anonymization of such data or the user's consent.
- Users must be able to withdraw their consent to the processing of traffic and location data at any time.
- The personalized service provider must inform the user of the type of data that will be processed, of the purposes and duration of the processing, and whether the data will be transmitted to a third party, prior to obtaining her consent.



European privacy laws

- Personal data obtained for different purposes may not be grouped.
- Usage data must be erased immediately after each session (except for very limited purposes).
- No fully automated individual decisions are allowed that produce legal effects concerning the data subject or significantly affect him and which are based solely on automated processing of data intended to evaluate certain personal aspects relating to them, such as their performance at work, creditworthiness, reliability, conduct, etc.

Industry and company regulations

- Industries and companies may have their own regulations
- An example, the U.S. Network Advertising Company Initiative
- prohibit the use of "personally identifiable information ("PII") [...] collected offline merged with PII collected online for online preference marketing unless the consumer has been afforded robust notice and choice about such merger before it occurs."

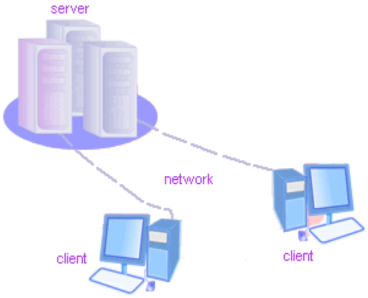
Principles of Fair Information Practices

- Basic principles to ensure privacy when dealing with personal data
  - Minimization: Collect minimum amount of data that is required, and store it for only as long as it is needed for the stated purpose
  - Consent: Enable users to opt-in and opt-out anytime (i.e. delete)
  - Openness: Make clear what data will be stored, for what and how long
  - Access: Enable users to inspect and correct their personal info
  - Accuracy: Ensure that data is correct and up-to-date, and it is propagated properly.
  - Security: Protect data against unauthorized access or modification

How personalized systems can satisfy these rules, or address privacy concerns?

- Pseudonymous users
  - Unidentifiable: impossible to track the real identity of pseudonymous users
  - Linkable for the personalized system: persistent identities across sessions
  - Unlinkable for third parties
  - Unobservable for third parties
- User data and user models are in servers, so server should be anonymized (pseudonymity infrastructure)
- Anonymization => disclosure => personalization ????

Client-side personalization



Personal data is stored in the client

- Advantages
  - Most data will be in the client – no privacy issues
  - Users might be inclined to disclose more info
- Challenges
  - How to implement collaborative algorithms?
  - How to protect confidential business rules that might be embedded in the personalization code?

Distribution, Encrypted Aggregation, Perturbation, and Obfuscation

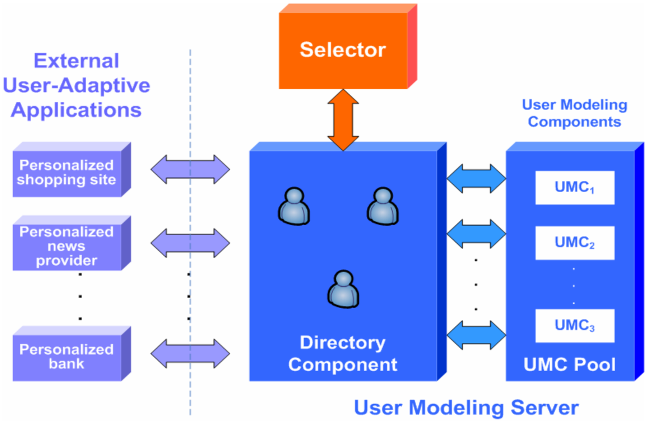
- Central repositories to support collaborative algorithms are very attractive for unauthorized access
- Mechanisms
  - Distribution:
    - distributed clusters with data of some users or P2P
  - Aggregation of Encrypted data:
    - encryption of rating in client side, and aggregation is done in some nodes without disclosing real ratings

Distribution, Encrypted Aggregation, Perturbation, and Obfuscation

- Mechanisms
  - Perturbation:
    - Central server receives ratings that have being systematically perturbed (e.g. adding a random number) before submission
    - 97% (90%) privacy => 13%(5%) recommendation error
  - Obfuscation:
    - A fraction of users' ratings is replaces by different values before submission to the central repository
    - Smallest impact: 90% of obfuscation => 5% error, but depends upon type of information, and level of ratings

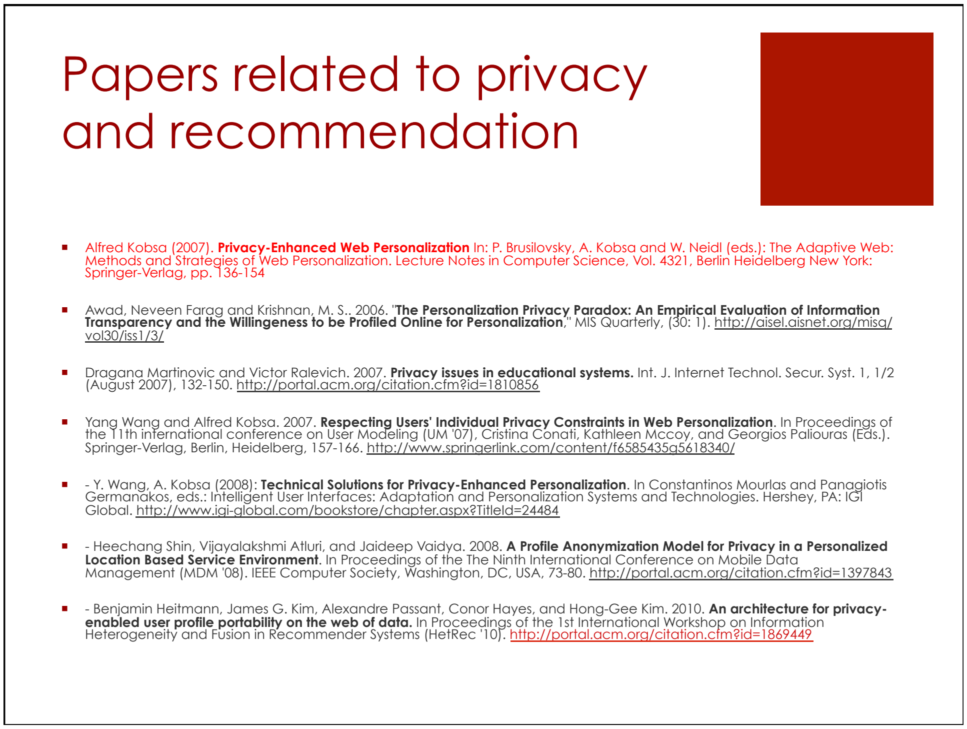
Personalizing Privacy

Adaptive selection of user models and components according privacy constraints



Conclusions

- User behavior is affected for their privacy concern
  - Users may disclose data if they see the benefits of doing so
- We need to address their privacy concerns
  - Several factors affect the willingness of providing information (e.g. trust)
- Privacy laws and agreements affect the implementation of personalized systems
  - As users we may give consent to many things without being aware of it because of complex privacy policies
- Mechanisms can deal with privacy risks.



Papers related to privacy and recommendation

- Alfred Kobsa (2007). **Privacy-Enhanced Web Personalization** In: P. Brusilovsky, A. Kobsa and W. Neidl (eds.): The Adaptive Web: Methods and Strategies of Web Personalization. Lecture Notes in Computer Science, Vol. 4321, Berlin Heidelberg New York: Springer-Verlag, pp. 136-154

- Awad, Neveen Farag and Krishnan, M. S.. 2006. "**The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingeness to be Profiled Online for Personalization**," MIS Quarterly, (30: 1). http://aisel.aisnet.org/misq/vol30/iss1/3/

- Dragana Martinovic and Victor Ralevich. 2007. **Privacy issues in educational systems.** Int. J. Internet Technol. Secur. Syst. 1, 1/2 (August 2007), 132-150. http://portal.acm.org/citation.cfm?id=1810856

- Yang Wang and Alfred Kobsa. 2007. **Respecting Users' Individual Privacy Constraints in Web Personalization**. In Proceedings of the 11th international conference on User Modeling (UM '07), Cristina Conati, Kathleen Mccoy, and Georgios Paliouras (Eds.). Springer-Verlag, Berlin, Heidelberg, 157-166. http://www.springerlink.com/content/f6585435g5618340/

- - Y. Wang, A. Kobsa (2008): **Technical Solutions for Privacy-Enhanced Personalization**. In Constantinos Mourlas and Panagiotis Germanakos, eds.: Intelligent User Interfaces: Adaptation and Personalization Systems and Technologies. Hershey, PA: IGI Global. http://www.igi-global.com/bookstore/chapter.aspx?TitleId=24484

- - Heechang Shin, Vijayalakshmi Atluri, and Jaideep Vaidya. 2008. **A Profile Anonymization Model for Privacy in a Personalized Location Based Service Environment**. In Proceedings of the The Ninth International Conference on Mobile Data Management (MDM '08). IEEE Computer Society, Washington, DC, USA, 73-80. http://portal.acm.org/citation.cfm?id=1397843

- - Benjamin Heitmann, James G. Kim, Alexandre Passant, Conor Hayes, and Hong-Gee Kim. 2010. **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). http://portal.acm.org/citation.cfm?id=1869449