




Personalization with User-Tailored Privacy

Alfred Kobsa

University of California, Irvine, CA
Microsoft Research, Redmond, WA



Examples for real-world personalization

New For You



The Man Who
Shot Liberty
Bells
Peter Lorre
DVD



~~\$29.95~~ **\$24.99**

[Why recommended?](#)

Google
繁體中文

Venus

Google 搜尋

好手氣

~~\$19.95~~ **\$14.99**

[Why recommended?](#)

[Why recommended?](#)

regarding product presentation and media types (e.g., text, graphics, video)

Personalization delivers benefits for both users and providers

Jupiter Communications, 1998: Personalization at 25 e-commerce sites boosted the number of new customers by 47% in the first year, and revenues by 52%.

Nielsen NetRatings, 1999:

- Registered visitors to portal sites spend over 3 times longer at their home portal than other users, and view 3 to 4 times more pages at their portal
- E-commerce sites offering personalized services convert significantly more visitors into buyers than those that don't.

Choicestream 2004 - 2009:

- 80% interested in personalized content
- 60% willing to spend a least 2 minutes answering questions about themselves

Tam & Hong, 2007: Users who received personalized music recommendations downloaded twice as much music, rated it higher, and browsed less for it

Eric Schmidt: 20-30% of Amazon purchases and 60% of Netflix views are a result of personalized recommendations




Many sources: Personalized ads enjoy significantly higher click-through rates

Downside of personalization

Personalized systems collect significantly more personal data than regular systems, and do this often in a very inconspicuous manner.

Many computer users are concerned about their privacy online

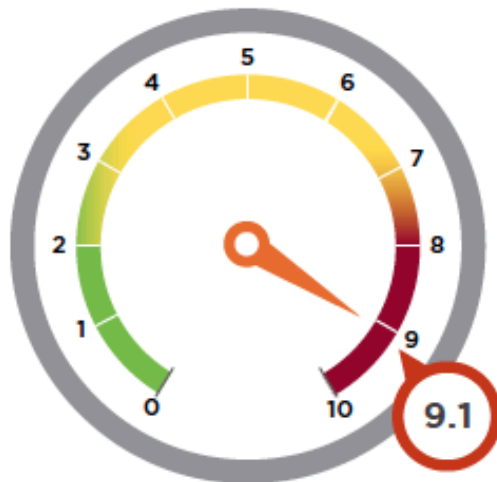
Number of users who reported:

- *being extremely or very concerned about **divulging** personal information online:*
67% (Forrester 1999), 74% (AARP 2000)
- *being (extremely) concerned about **being tracked** online:*
77% (AARP 2000)
- ***leaving** web sites that required registration information:*
41% (Boston Consulting 1997)
- *having entered **fake** registration information:*
40% (GUV 1998), 27% (Boston Consulting 1997), 32% (Forrester 1999)
- *having **refrained from shopping** online due to privacy concerns, or **bought less**:*
32% (Forrester 1999), 32%  35%  54%  IBM 1999, 24% (AARP 2000)
- *wanting internet sites **ask for permission** to use personal data: 81% (Pew 2000)*
- *being willing to give out personal data for getting something **valuable in return**:*
31% (GUV 1998), 30% (Forrester 99), 51% (Personalization Consortium)

2012 Privacy attitudes in the U.S.



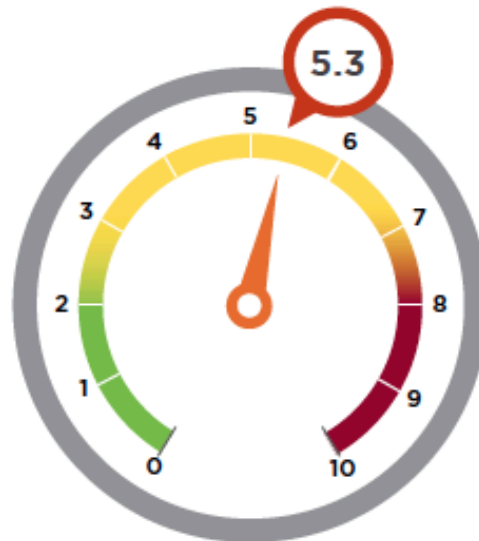
CONSUMER CONCERN



Consumer concern is extremely high this quarter; 91 percent of U.S. adults worry about their privacy online.



CONSUMER MISTRUST



Consumer mistrust is high this quarter; 53 percent of U.S. adults do not completely trust businesses with their personal information.



BUSINESS IMPACT

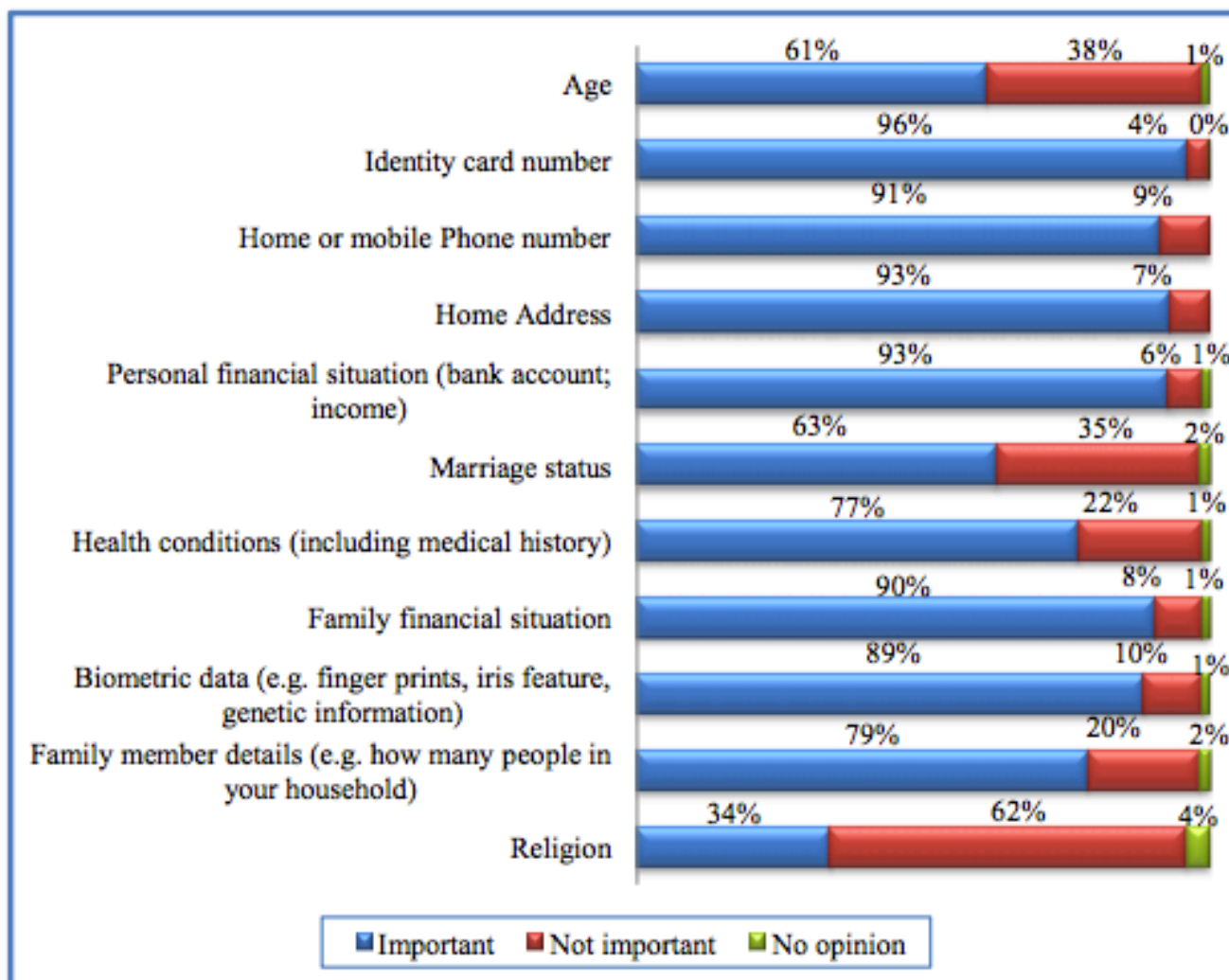


Business impact is extremely high this quarter; 88 percent of U.S. adults say they avoid companies that do not protect their privacy.

Source: Q2 2012 TRUSTe Privacy Index

2011 Privacy attitudes in Hong Kong

Chart 1: Distributions of privacy importance of personal data by type (%)



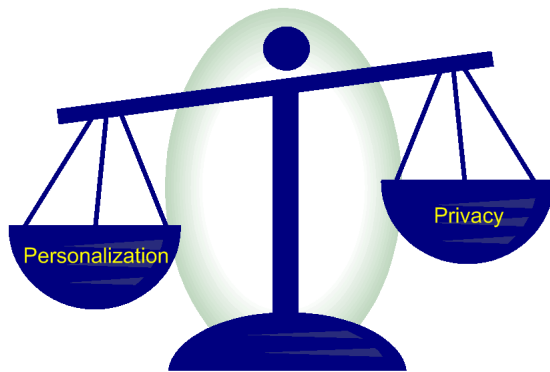
Source: Policy 21 Ltd., Feb. 2012

Either Personalization or Privacy?

Balancing Privacy with Personalization

Privacy vs. Personalization: A Delicate Balance

- Personal data of users are indispensable for personalized interaction
- Users are reluctant to give out personal data



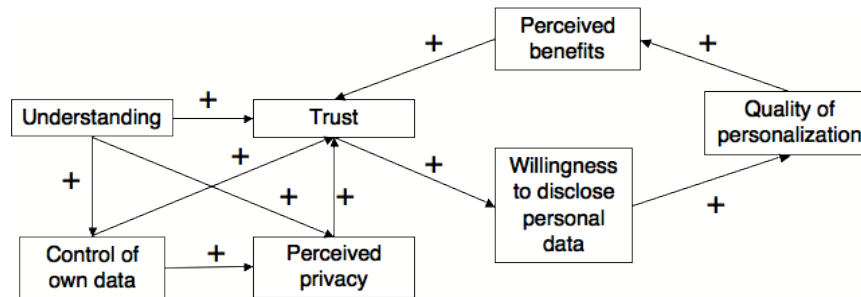
Tradeoff between privacy and personalization?

[print this article](#) | [e-mail a colleague](#)

Personalization Vs. Privacy Debate Heating Up

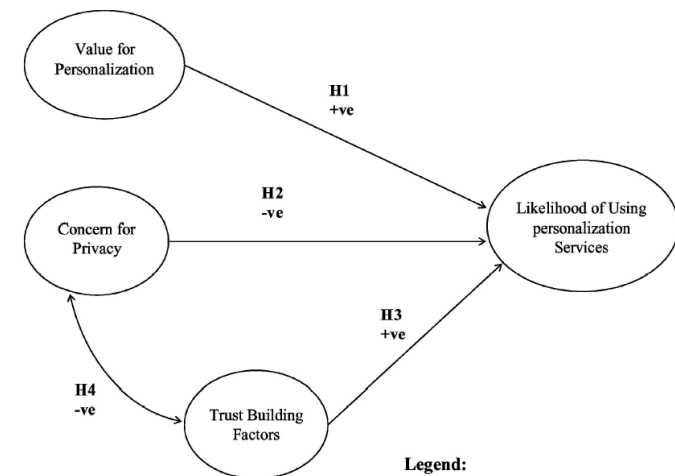
>>> ClickZ News

The tension between privacy and personalization is more complex than that...



- Indirect relationship between privacy and personalization
- Situation-dependent
- Many mitigating factors

People use “privacy calculus” to decide whether or not to disclose personal data, e.g. for personalization purposes



Privacy-Enhanced Personalization

Can we have good personalization and good privacy at the same time?

How can personalized systems maximize their personalization benefits, while at the same time being compliant with the privacy constraints that are in effect?



What are those privacy constraints, and how do we deal with them?

Privacy constraints

- A. Users' individual privacy preferences
(and factors that influence them in a given situation)
- B. Privacy norms (laws, self-regulation, principles)

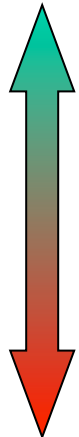
Reconciliation of privacy and personalization

1. Use of privacy-enhancing technology
2. Privacy-minded user interaction design

Individual privacy preferences for personal data

Influenced by...

Information type



- Basic demographic and lifestyle information, personal tastes, hobbies
- Internet behavior and purchases
- Extended demographic information
- Financial and contact information
- Credit card and social security numbers

Data values

- Willingness to disclose certain data decreases with deviance from group average
(Confirmed for age, weight, salary, spousal salary, credit rating and amount of savings)














Privacy norms

- Privacy laws
More than 50 countries and 100 states worldwide (e.g., Macau, Hong Kong)
- Industry self-regulation
Company-internal, industry sector specific (NAI), nation-wide (Singapore)
- Industry self-regulation with government enforcement **NEW!**
Proposed U.S. Consumer Privacy Bill of Rights
- Privacy principles
 - supra-national (OECD, APEC)
 - national (Australia, Canada, New Zealand...)
 - member organizations (ACM)

Quite a few privacy norms...

- **require explicit user *consent* before personal data may be collected**
- **forbid a number of frequently used personalization methods unless the user *consents***

Privacy laws and regulations restrict the permissibility of personalization methods

-  Usage logs must be deleted after each session 
-  Usage logs of different services may not be combined (except for accounting purposes) 
-  User profiles are permissible only if pseudonyms are used. (Profiles retrievable under pseudonyms shall not be combined with data relating to the bearer of the pseudonym.) 
-  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 him, such as his performance at work, creditworthiness, reliability, conduct, etc. 

-  Anonymous or pseudonymous access and payment must be offered if technically possible and reasonable. 
-  Users must be able to withdraw their consent on processing traffic or location data at any time 

Privacy constraints, and how to deal with them

Privacy constraints

- A. Users' individual privacy preferences in a given situation
(and factors that influence them)
- B. Privacy norms (laws, self-regulation, principles)

Reconciliation of privacy and personalization

1. Use of privacy-enhancing technology
2. Privacy-minded user interaction design

1. Enabling Websites to Respect Privacy Norms and User Preferences in Web Personalization

- Internationally operating personalized website are often obliged to cater to different national privacy laws, even if they are not located in the respective jurisdiction.
- They also need to take users' individual privacy preferences and industry self-regulation into account.
- International privacy laws and user privacy preferences often disallow the use of various personalization methods

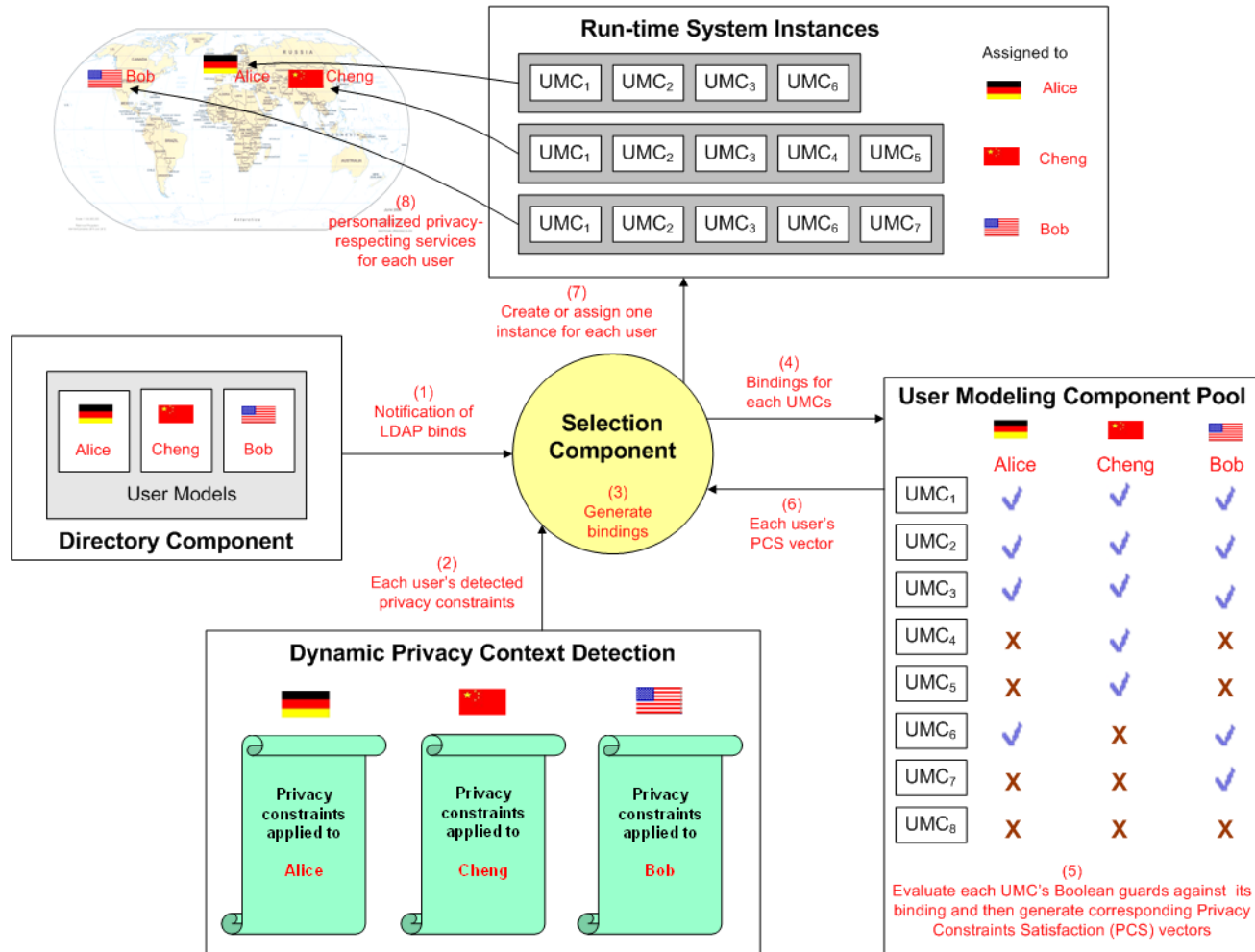
Our approach

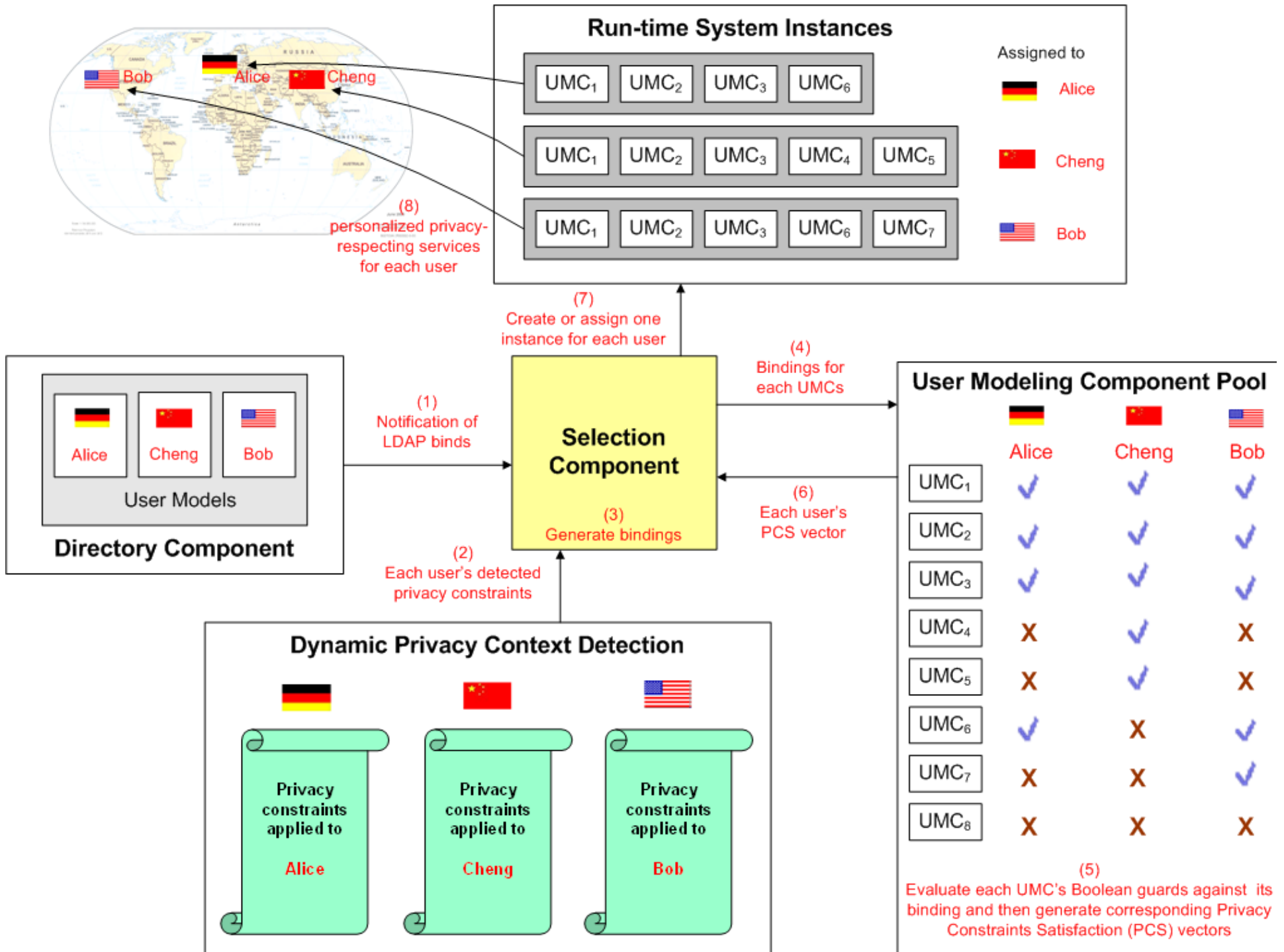
Develop a mechanism that dynamically selects those user modeling methods that *comply with the currently prevailing privacy constraints*, namely

- the user's individual privacy preferences
- the privacy norms that apply to the user

 “User-tailored privacy”

Ex: Internationally operating recommender that adapts to privacy constraints





The privacy constraints



Privacy constraints applied to **Alice**

German Tele-Service Data Protection Law

Section 4(2)-4(4): profiling

Combining user profiles retrievable under pseudonyms with data relating to the bearer of the pseudonym, is prohibited.

Personal data to be erased immediately after each session except for very limited purposes.

.

.

.



Privacy constraints applied to **Cheng**

Cheng's own privacy preferences:

"Dislike being tracked"

.

.

.



Privacy constraints applied to **Bob**

Network Advertising Initiative (NAI) Self-Regulatory Principals

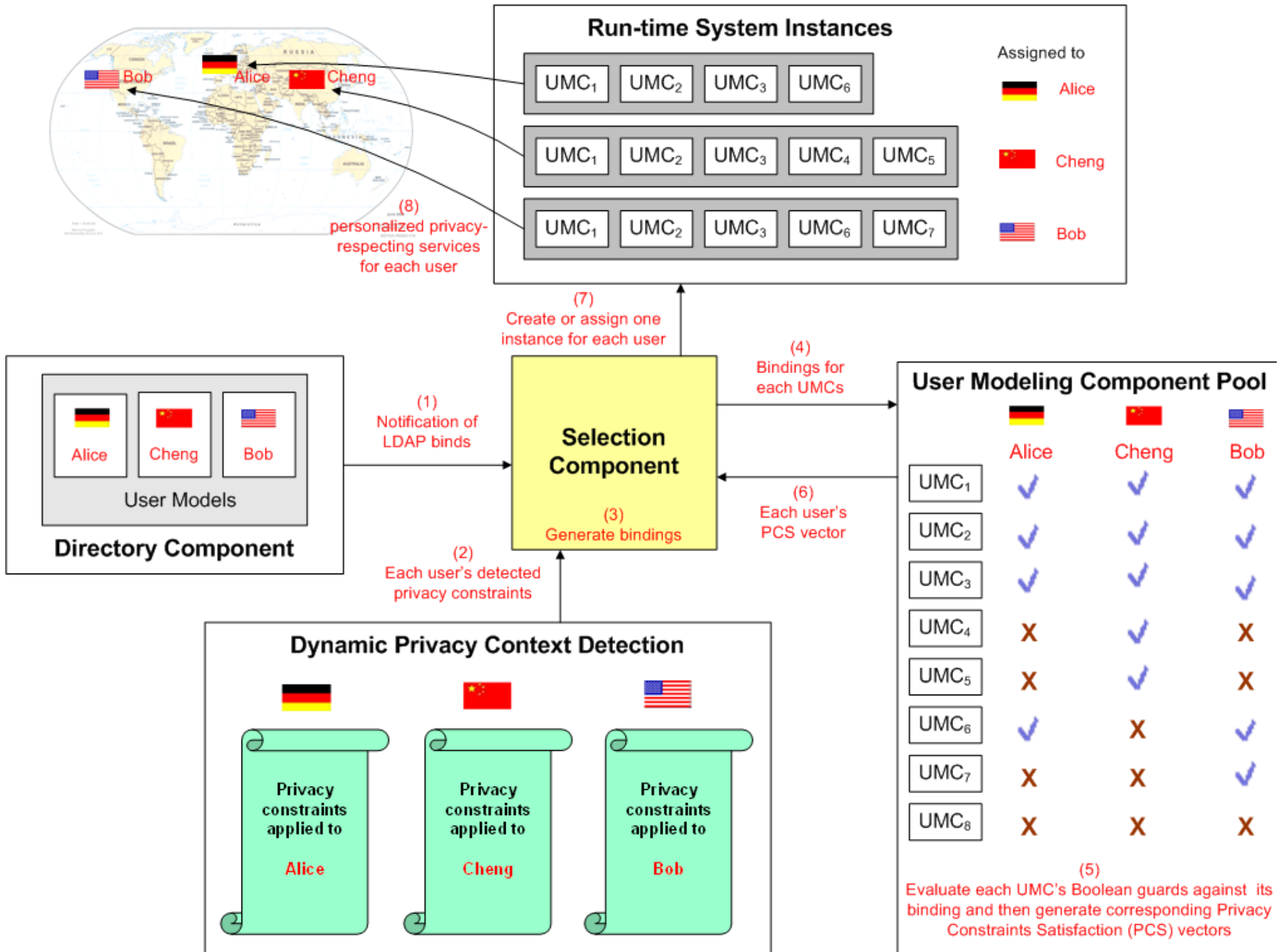
Section II: NAI's Statement of Purposes

Merging non-personally identifiable use data with personally identifiable demographic data, is prohibited unless user give prior affirmative consent.

.

.

.



Resource requirements for google.com (April 2011)

- No. 1 website in terms of traffic
- 3.24 billion visits per month
- 1250 visits per second
- cloud of 2,500 nodes needed to host our personalized privacy architecture
- Google currently uses several 100,000 servers



User evaluation

BOOK CHOICES

You currently have a choice of **1,000,000** books.

BROWSE

Data Protection
Your personal data is protected by us

Personalization
We want to offer you a personalized service

Security
Your security is our top priority

Affiliation
Earn money with your website

Buy & Figures
Payment by Amazon.com

Gift Services
This makes gifts even more fun!

Reviews
Write reviews to win exciting prizes!

Amazon.com on



You can change your privacy preferences anytime in the privacy control panel on the right-hand side.

1.) Please enter a login name (your name or a pseudonym)

- Login name:
- No answer

2.) How old are you?

- 18-20
- 21-25
- 26-30
- 31-35
- 36-40
- 41-50
- 51-60
- >60
- No answer

3.) What is your occupation / degree program?

- Occupation / degree program:
- No answer

4.) What are your hobbies? (Check all that apply.)

- Sport
- Music
- Model making
- Computers

PRIVACY CONTROL

Your privacy preferences determine how we personalize book selections for you

1. Your Privacy Preferences

Check any item that you allow:

- Use your data for other purposes
- Keep your usage data longer
- Track what you do on our site
- Use your location data
- Merge your usage and identity data

2. How do we personalize then?

Legend: use, not use

- Clustering
- Rule-based reasoning I
- Rule-based reasoning II
- Incremental learning
- One-time learning I
- One-time learning II

Experimental Procedures (partly based on deception)

1. Instructions to subjects

- “Usability test with new version of a well-known online book retailer”
- Answering questions to allegedly obtain better book recommendations
- No obligation to answer any question, but helpful for better recommendation.
- Data that subjects entered would purportedly be available to company
- Possibility to buy *one* of the recommended books with a 70% discount.
- Reminder that if they buy a book, ID card and credit card would be checked (subjects were instructed beforehand to bring these documents if they wish to buy)

2. Answering interest questions in order to “filter the selection set” (anonymous)

- 32 questions with 86/64 answer options become presented (some free-text)
- Most questions were about users’ interests (a very few were fairly sensitive)
- Answering questions decreased the “selection counter” in a systematic manner
- After nine pages of data entry, users are encouraged to review their entries, and then to view the recommended books that purportedly match their interests

BOOK CHOICES

You currently have a choice of **1,000,000** books.

BROWSE

Data Protection
Your personal data is protected by us

Personalization
We want to offer you a personalized service

Security
Your security is our top priority

Affiliation
Earn money with your website

Buy & Figures
Payment by Amazon.com

Gift Services
This makes gifts even more fun!

Reviews
Write reviews to win exciting prizes!

Amazon.com on



You can change your privacy preferences anytime in the privacy control panel on the right-hand side.

1.) Please enter a login name (your name or a pseudonym)

- Login name:
- No answer

2.) How old are you?

- 18-20
- 21-25
- 26-30
- 31-35
- 36-40
- 41-50
- 51-60
- >60
- No answer

3.) What is your occupation / degree program?

- Occupation / degree program:
- No answer

4.) What are your hobbies? (Check all that apply.)

- Sport
- Music
- Model making
- Computers

PRIVACY CONTROL

Your privacy preferences determine how we personalize book selections for you

1. Your Privacy Preferences

Check any item that you allow:

- Use your data for other purposes
- Keep your usage data longer
- Track what you do on our site
- Use your location data
- Merge your usage and identity data

2. How do we personalize then?

Legend: use, not use

- Clustering
- Rule-based reasoning I
- Rule-based reasoning II
- Incremental learning
- One-time learning I
- One-time learning II

Experimental Procedures (partly based on deception)

1. Instructions to subjects

- “Usability test with new version of a well-known online book retailer”
- Answering questions to allegedly obtain better book recommendations
- No obligation to answer any question, but helpful for better recommendation.
- Data that subjects entered would purportedly be available to company
- Possibility to buy *one* of the recommended books with a 70% discount.
- Reminder that if they buy a book, ID card and credit card would be checked (subjects were instructed beforehand to bring these documents if they wish to buy)

2. Answering interest questions in order to “filter the selection set” (anonymous)

- 32 questions with 86/64 answer options become presented (some free-text)
- Most questions were about users’ interests (a very few were fairly sensitive)
- Answering questions decreased the “selection counter” in a systematic manner
- After nine pages of data entry, users are encouraged to review their entries, and then to view the recommended books that purportedly match their interests

Experimental Procedures (cont' d)

3. “Recommendation” of 50 books (anonymous)

- 50 predetermined and invariant books are displayed (popular fiction, politics, travel, sex and health advisories)
- Selected based on their low price and their presumable attractiveness for students
- Prices of all books are visibly marked down by 70%, resulting in out-of-pocket expenses between \$2 and \$12 for a book purchase.
- Extensive information on every book available

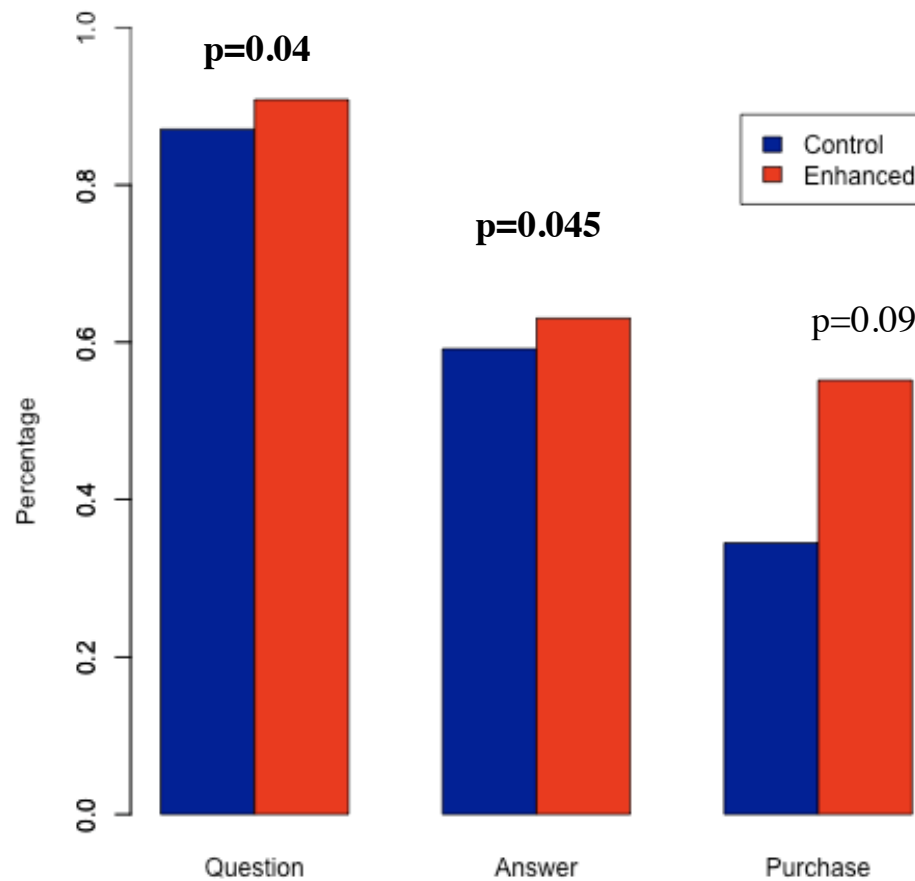
4. Optional purchase of one book (identified)

- Subjects may purchase one book if they wish
- Those who do are asked for their names, shipping and credit card data.

5. Completing questionnaires

6. Verification of name, address and CC data (if book bought)

Results: disclosure and purchases



Downsides of “informed consent”, “transparency & control”

- **Can become unwieldy**
 - Facebook has “labyrinthian” controls (U.S. Consumer Magazine)
- **Presumes that users are rational decision makers**
 - Herding and order effect (Acquisti et al. 2011)
 - Privacy information raises privacy fears (Knijnenburg et al. 2012)
 - If misplaced in the workflow, privacy notices become ignored (Egelman et al. 2009)
 - Professionalism of UI design matters (John et al. 2011)
 - It matters what the default is (Lai & Hui 2006)
 - Control may lead to over-disclosure (Brandlmarte et al. 2012)



Is informed consent a slight of hands?

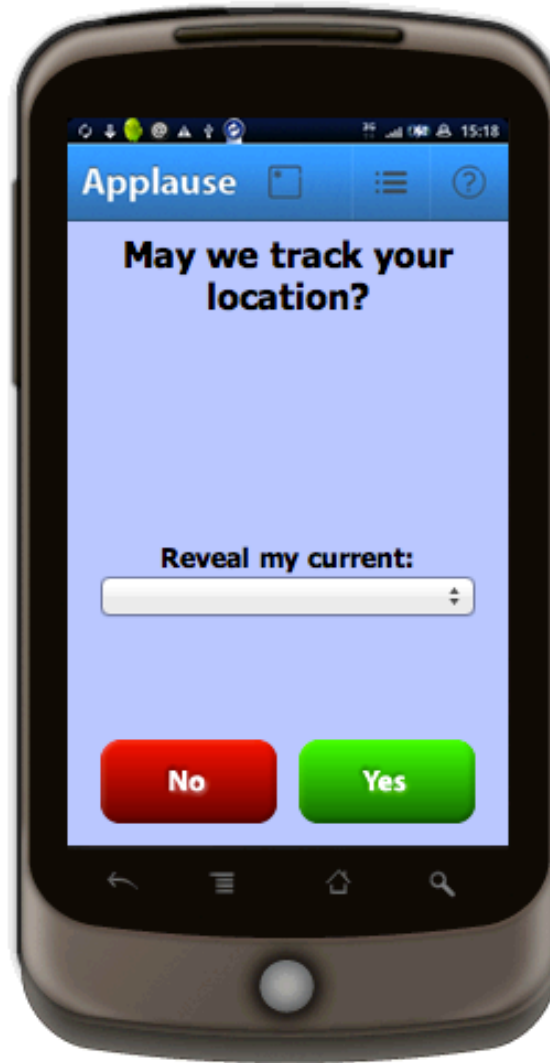
Privacy interventions

“Privacy nudging”

- Appropriate defaults
- Reminders
- Rewards

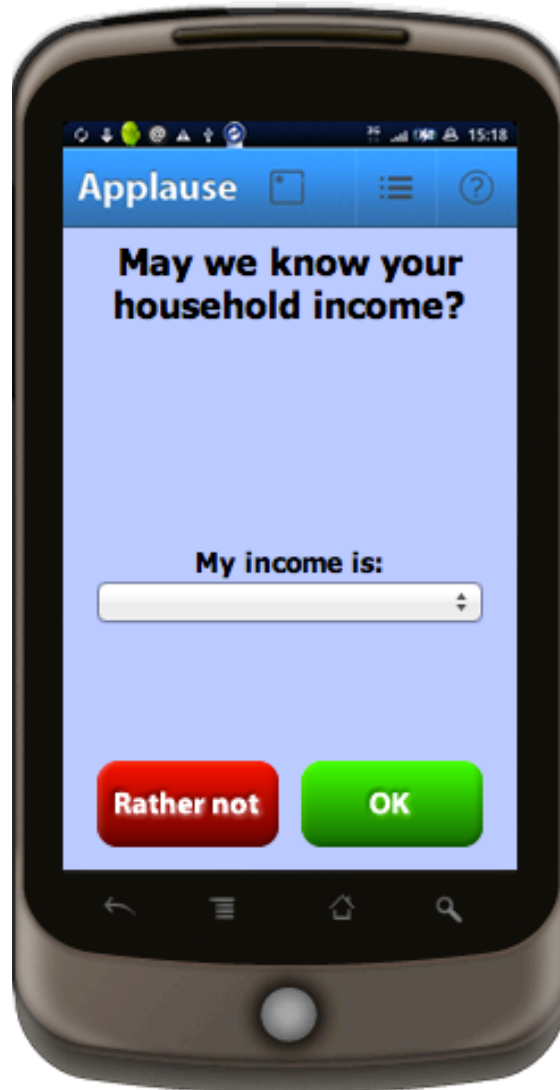
Problem: one-size-fits-all approach

Request for permission



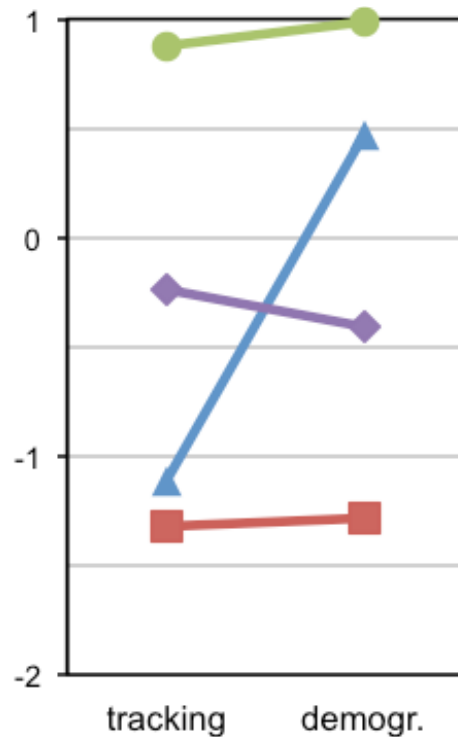
Funded by Samsung, Ericsson

Disclosure request

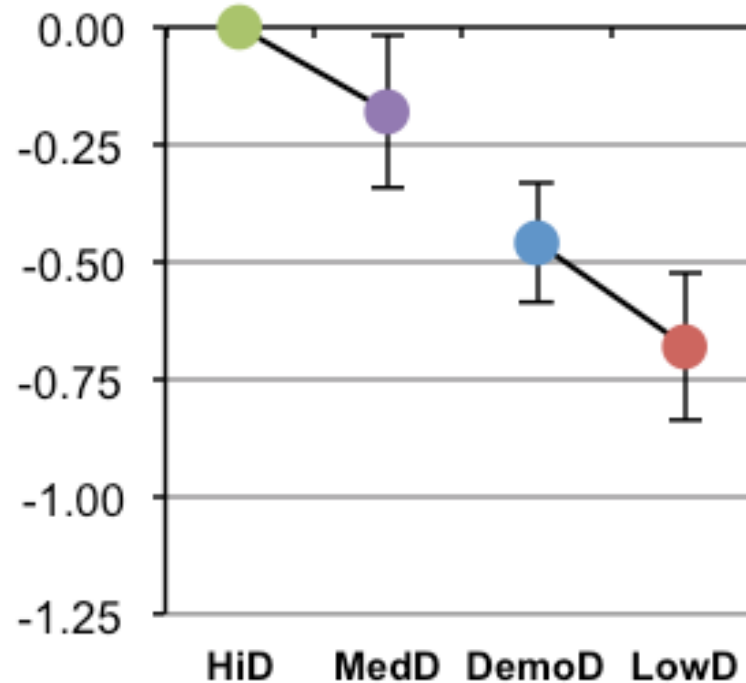


User clusters based on the disclosure of mobile tracking and demographic data

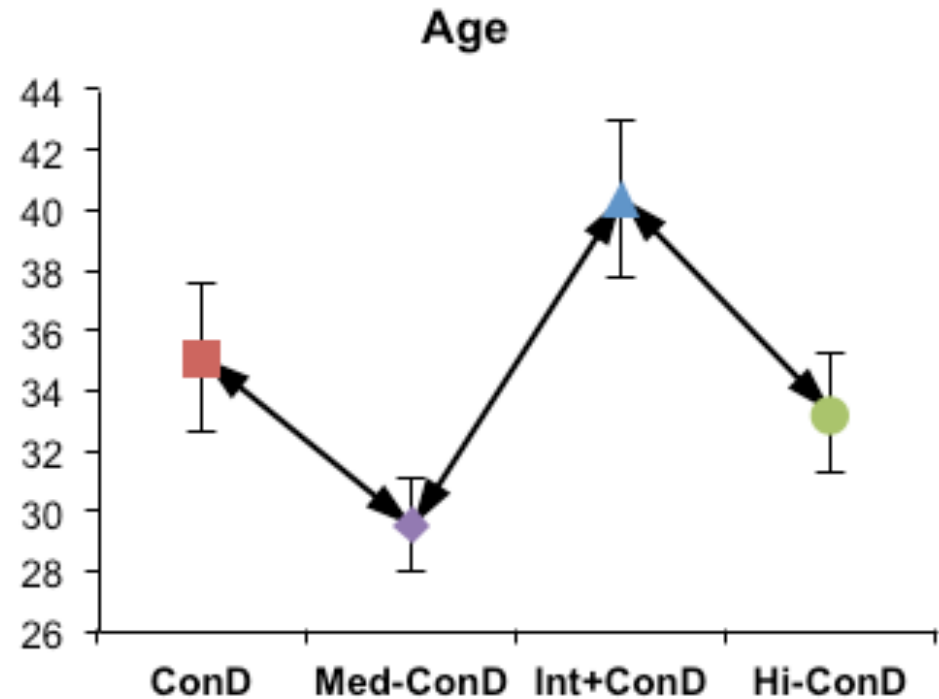
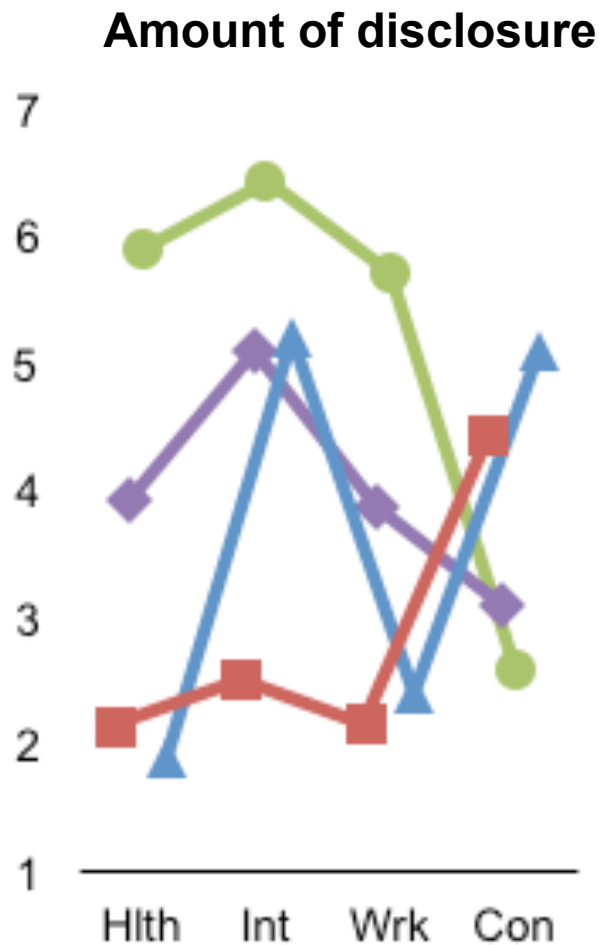
Amount of disclosure



Mobile Internet usage



User clusters based on the disclosure of four types of data to an online retailer



Solution: personalized privacy

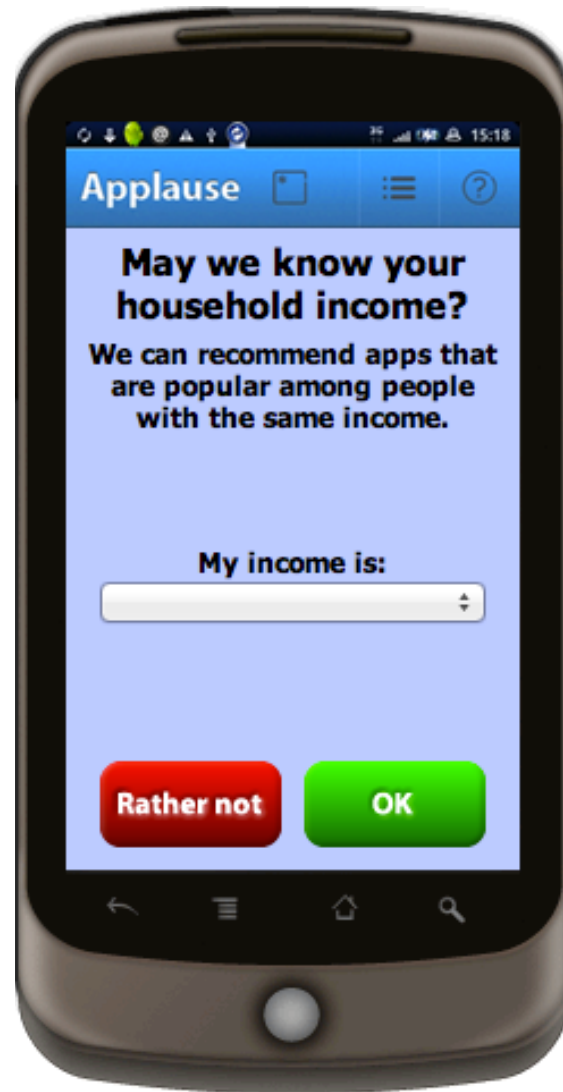
At development time:

Run user studies and identify groups with different disclosure behaviors, and other characteristics of these groups (age, gender, internet usage).

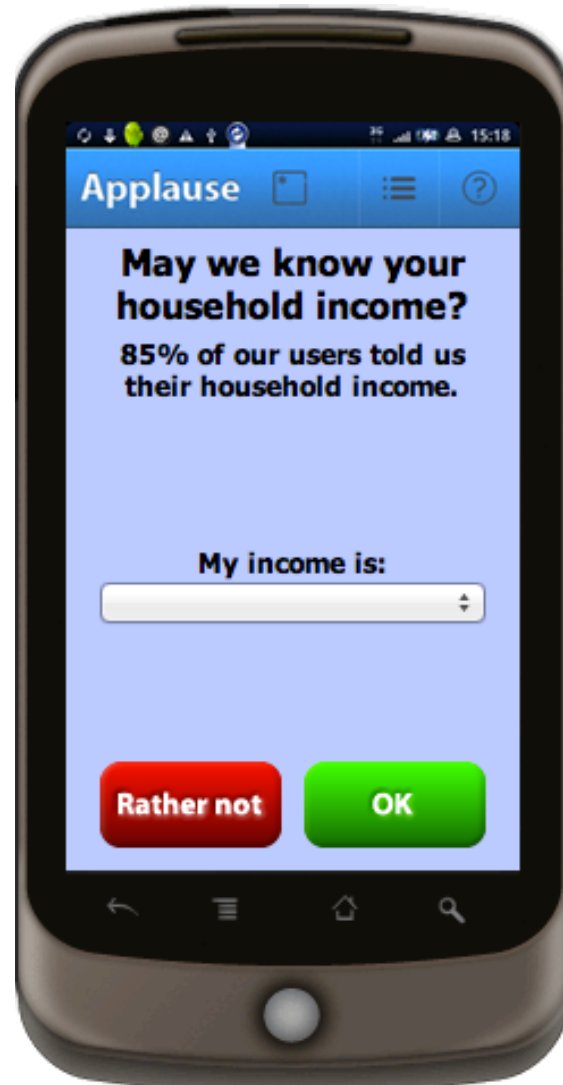
At runtime:

1. Determine a user's characteristics (age, gender,)
2. Predict the user's privacy behavior based thereon
3. Cater to this anticipated behavior
 - Set default privacy preferences for the user
 - Adjust privacy-related information

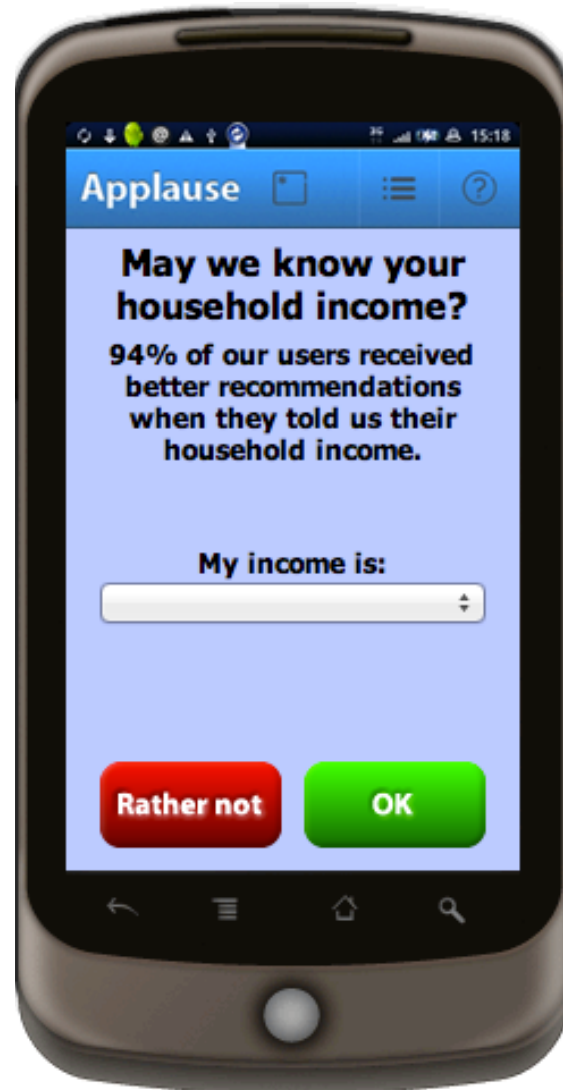
Disclosure request – Explanation of use



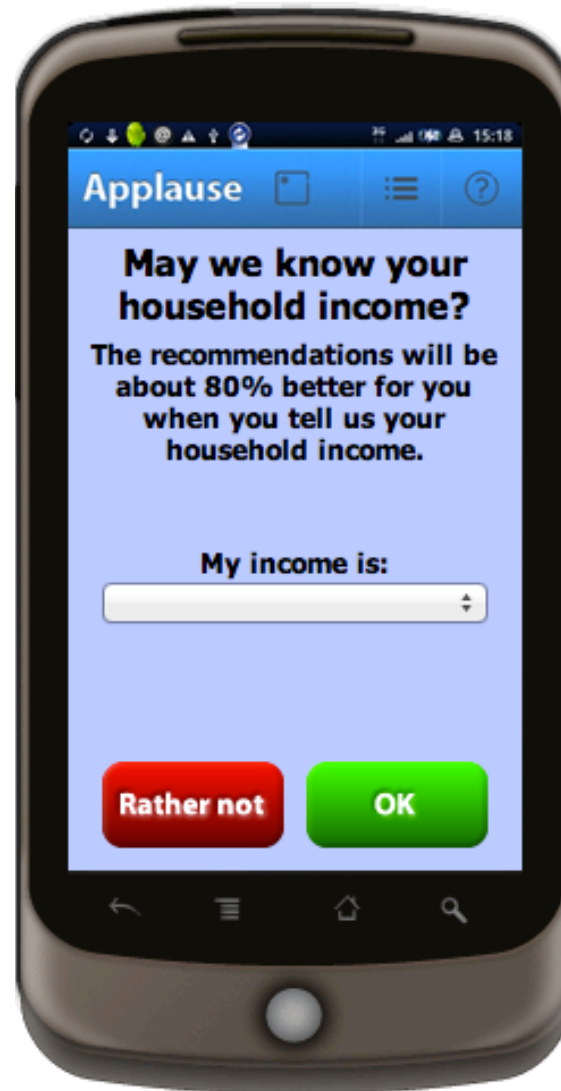
Disclosure request – Social cues



Disclosure request – Past benefit for others



Disclosure request – Projected benefit for me



Adjusting privacy-related system behavior once user has been classified

| Best strategies for MALES with LOW disclosure tendency | |
|---|---|
| <i>Goal</i> | <i>Best strategy</i> |
| High demographics disclosure | Demographics first, 'explanation' justification. |
| High context data disclosure | Context first, no justification. |
| High overall disclosure | Context first, 'useful for you' justification. |
| High satisfaction | Context first, 'useful for others' justification or demographics first, 'useful for you' justification. |
| All of the above | Demographics first, no justification. |
| Best strategies for FEMALES with LOW disclosure tendency | |
| <i>Goal</i> | <i>Best strategy</i> |
| High demographics disclosure | Demographics first, 'number of others' justification. |
| High context data disclosure | Context first, 'useful for you' justification. |
| High overall disclosure | Demographics first, 'explanation' justification. |
| High satisfaction | Context first, 'useful for you' justification. |
| All of the above | Demographics first, 'explanation' justification. |

There is no magic bullet for reconciling personalization with privacy



Effort is comparable to

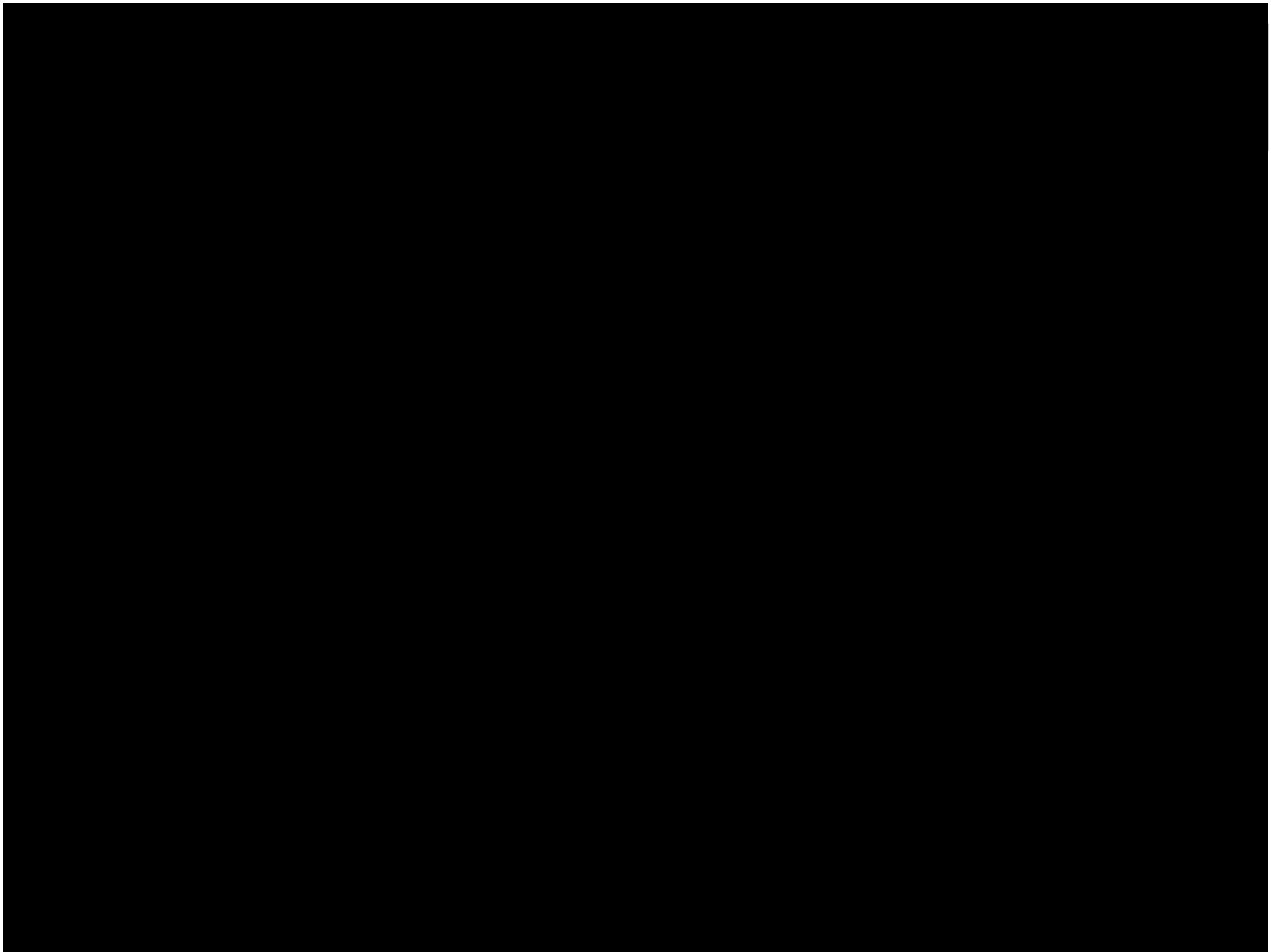
- ... making systems secure
- ... making systems fast
- ... making systems reliable

Privacy-Enhanced Personalization: Process approach needed

1. Gain the user's trust

- Respect the user's privacy attitude (and let the user know)
 - Respect privacy laws / industry privacy agreements
- Provide benefits (including optimal personalization within the given privacy constraints)
- Increase the user's understanding (don't do magic)
- Give users control
- Use trust-enhancing methods
- Use privacy-enhancing technology (and let the user know)

2. *Then be patient, and most users will incrementally come forward with personal data / permissions if the usage purpose for the data and the ensuing benefits are clear and valuable enough to them.*



Survey with system mockup

- For theory construction
 - 200 subjects via Amazon Turk (U.S. only)
 - 56 through Craigslist.com
- For theory confirmation
 - 239 participants via Amazon Turk
- 223 males, 266 females, 2 n.a.
- Ages from 18 to 60+, median 25-30

