Expecting the Unexpected
Understanding Mismatched Privacy Expectations Online

Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, Ruogu Kang

Carnegie Mellon University

SOUPS 2016 | June 22, 2016
What information does this website collect?
Bank of America U.S. Online Privacy Notice

Last updated July 24, 2014

This U.S. Online Privacy Notice (Notice) applies to this Bank of America online interface (i.e., website or mobile application) and any Bank of America U.S. affiliate or subsidiary online interface that links to this Notice, (each, a “Site,” and, collectively, “Sites”). The term “Bank of America” or “we” or “us” or “our” in this Notice refers to banking and non-banking U.S. affiliates or subsidiaries of Bank of America Corporation that link to this Notice. This Notice describes how Sites may collect, use and share information from or about you, and explains how information may be collected and used for advertising purposes.

Bank of America provides other online interfaces not covered by this Notice. If you visit or access your accounts from one of these sites, please review the online privacy practices of that site to understand how your online information may be collected, used and shared.

For U.S. account holders and visitors to this Site, we will use and share any information that we collect from or about you in accordance with the Bank of America U.S. Consumer Privacy Notice, which provides choices in the use and sharing of information. For Non-U.S. account holders utilizing this Site, we will use and share your account information in accordance with the privacy disclosure that covers your account and with the privacy and security rules applicable to the Bank of America affiliate or subsidiary that provides that account to you.

Additional information on our Privacy & Security practices may be found on our Sites and within Frequently Asked Questions (FAQ). Although the additional information is provided as a resource, the terms and conditions of this Notice control, and by using the Site, you agree to the terms and conditions of this Notice.

Collecting and Using Information

Personal Information We Collect Online

Personal Information means personally identifiable information such as information you provide via forms, surveys, applications or other online fields including name, postal or email addresses, telephone, fax or mobile numbers, or account numbers.

How We Use Personal Information

We may use Personal Information:

- To process transactions and provide you products and services.
- To verify your identity or to prevent or detect fraud.
- To meet legal and regulatory requirements.
- To contact you about your accounts.
- To manage relationships with you, including by providing information about products and services.
- To provide you offers, promotions, and materials from Bank of America and third parties.
- To conduct business and operational activities.
- To evaluate the performance of our products and services.
- To protect against and prevent fraud, unauthorized access or use, and other malicious activity.
- To improve our websites and mobile applications.

We may share Personal Information with third parties to the extent permitted by law, as described in our Privacy Disclosure. You may review our Privacy Disclosure by clicking on the “Privacy” link in the Site's footer.

How to Contact Us

If you have any questions about this Notice or privacy-related concerns, you may contact us by writing to:

Bank of America
Consumer Relations
P.O. Box 95407
Miami, FL 33197-4075

You may also contact us by calling 1-800-432-1000.

Privacy Notice is subject to change. Please check our web site for the most recent version of the Privacy Notice.
Protecting individual health information

To the extent that we receive, maintain, or process an individual’s protected health information, Bank of America may disclose that information as authorized by and in accordance with applicable federal and/or state law.
unexpected & surprising practices easily overlooked among practices that are expected or irrelevant for the use context.
“the question is not whether consumers should be given a say over unexpected uses of their data; rather, the question is how to provide simplified notice and choice.”

Edith Ramirez
FTC Chairwoman
January 2015
research questions

What practices are expected or unexpected?

How can we measure expectations and mismatches in expectations?

How can we emphasize unexpected practices in privacy notices?
types of expectations

Privacy literature

Privacy preferences
Willingness to share/disclose
Desired level of privacy
Actual privacy

malleable, uncertain, context-dependent

Altman. The environment and social behavior: Privacy, personal space, territory, and crowding. 1975.
Other domains, e.g. consumer psychology, distinguish different types of expectations

Miller:
- Ideal (what it could be)
- Expected (what it likely will be)
- Deserved (what it should be)
- Minimum Tolerable (what it must be)
privacy expectations

privacy expectations
(what is likely)

VS

privacy preferences
(what it should be)

actual practices
elicit privacy expectations
• present participants with actual websites in online study
• ask participants to rate likelihood that website engages in certain data practices (objective expectation)

privacy policy analysis
• extract practices disclosed in website privacy policies

identify mismatches
• compare likelihood expectations with disclosed practices
data practices considered

**data collection**
- 4 information types: contact, financial, health, current location
- 2 scenarios: user with account, user without account

**sharing with third parties**
- 4 information types: contact, financial, health, current location
- 2 purposes: sharing for core purpose / other purpose

**data deletion**
- does the website allow deletion of personal data?
website features

**website type:**
- finance
- health
- dictionary

**popularity:**
- high rank
- low rank

**ownership:**
- private
- government

user features

**website experience**
- recent use, has account, familiarity, trust

**demographics**
- age, gender, education, occupation, computer background

**privacy**
- privacy protective behavior
- privacy knowledge
- negative online experience
- online privacy concern (IUIPC)
study deployment

between-subjects study

• 16 websites
• 240 participants (Amazon Mechanical Turk)

• each participant randomly assigned to one website; 15 participants per website
“Imagine that you are browsing [website name] website. You do not have a user account on [website name], that is, you have not registered or created an account on the website”

“What is the likelihood that [website name] would collect your information in this scenario? …”

<table>
<thead>
<tr>
<th>Collects your Contact information</th>
<th>Likely</th>
<th>Somewhat likely</th>
<th>Somewhat unlikely</th>
<th>Unlikely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email address</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Postal address</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Phone number</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Other</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Please specify</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
privacy policy analysis

ways to extract data practice statements from privacy policies

• machine-readable policy specification (e.g. P3P)
• (semi-)automated extraction of data practices from policy
• manual annotation by experts
privacy policy analysis

extracting data practice statements from privacy policies

• **manual annotation by experts**
  • Yes website engages in practice
  • No website does not engage in practice
  • Unclear not clear if website engages in practice
  • **Not addressed** the policy is silent regarding practice

• 2 annotators analyzed 16 websites’ privacy policies
privacy policy analysis
types of mismatched expectations

- yes (website) vs no (user)
  - website shares data, but user doesn’t expect it
  - user may give up data unknowingly; website may lose trust

- no (website) vs yes (user)
  - website doesn’t share data, but user thinks so
  - user may not use website & lose utility; website may lose customer

Different types of mismatches may impact user privacy differently
results

expectations

mismatches with policy statements
impact of website characteristics

• participants expect almost all websites to collect location and contact information and share it for core purposes

• website type had statistically significant effect on participants’ expectations
  – for collection of financial & health information
  – for sharing of financial and health information

• popularity and ownership had no effect
impact of user characteristics

**collection of location information with account**
- recent use: NO
- privacy concern: YES

**collection of health information without account**
- privacy knowledge: NO

**sharing of location information for core purpose**
- trust in website: YES
- privacy concern: YES

**sharing of contact information for core purpose**
- recent use: NO
- privacy concern: YES

**sharing of financial information for other purposes**
- trust in website: NO

**sharing of health information for other purposes**
- trust in website: YES

**allow deletion of personal data**
- age: NO
- recent use: NO
- trust in website: YES
mismatched expectations

<table>
<thead>
<tr>
<th>Category</th>
<th>Explicit match (NN,YY)</th>
<th>Explicit mismatch (NY,YN)</th>
<th>Unclear mismatch (UY, UN)</th>
<th>NA mismatch (NaY, NaN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td>33%</td>
<td>54%</td>
<td>13%</td>
<td>0%</td>
</tr>
<tr>
<td>Financial</td>
<td>20%</td>
<td>43%</td>
<td>12%</td>
<td>25%</td>
</tr>
<tr>
<td>Health</td>
<td>19%</td>
<td>19%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Location</td>
<td>36%</td>
<td>8%</td>
<td>37%</td>
<td>19%</td>
</tr>
</tbody>
</table>

Collect without account:
- Explicit match (NN,YY): 33%
- Explicit mismatch (NY,YN): 54%
- Unclear mismatch (UY, UN): 13%
- NA mismatch (NaY, NaN): 0%

Collect with account:
- Explicit match (NN,YY): 89%
- Explicit mismatch (NY,YN): 5%
- Unclear mismatch (UY, UN): 6%
- NA mismatch (NaY, NaN): 0%

Share for core purpose:
- Explicit match (NN,YY): 70%
- Explicit mismatch (NY,YN): 12%
- Unclear mismatch (UY, UN): 19%
- NA mismatch (NaY, NaN): 0%

Share for other purpose:
- Explicit match (NN,YY): 23%
- Explicit mismatch (NY,YN): 65%
- Unclear mismatch (UY, UN): 13%
- NA mismatch (NaY, NaN): 0%
mismatched expectations

<table>
<thead>
<tr>
<th>Category</th>
<th>Collect without account</th>
<th>Collect with account</th>
<th>Share for core purpose</th>
<th>Share for other purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit match (NN,YY)</td>
<td>33%</td>
<td>89%</td>
<td>70%</td>
<td>23%</td>
</tr>
<tr>
<td>Explicit mismatch (NY,YN)</td>
<td>54%</td>
<td>5%</td>
<td>12%</td>
<td>65%</td>
</tr>
<tr>
<td>Unclear mismatch (UY, UN)</td>
<td>13%</td>
<td>6%</td>
<td>19%</td>
<td>13%</td>
</tr>
<tr>
<td>NA mismatch (NaY, NaN)</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Financial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit match (NN,YY)</td>
<td>20%</td>
<td>46%</td>
<td>41%</td>
<td>44%</td>
</tr>
<tr>
<td>Explicit mismatch (NY,YN)</td>
<td>43%</td>
<td>17%</td>
<td>28%</td>
<td>25%</td>
</tr>
<tr>
<td>Unclear mismatch (UY, UN)</td>
<td>12%</td>
<td>12%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>NA mismatch (NaY, NaN)</td>
<td>25%</td>
<td>25%</td>
<td>31%</td>
<td>31%</td>
</tr>
<tr>
<td>Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit match (NN,YY)</td>
<td>19%</td>
<td>18%</td>
<td>12%</td>
<td>20%</td>
</tr>
<tr>
<td>Explicit mismatch (NY,YN)</td>
<td>19%</td>
<td>20%</td>
<td>13%</td>
<td>11%</td>
</tr>
<tr>
<td>Unclear mismatch (UY, UN)</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>NA mismatch (NaY, NaN)</td>
<td>63%</td>
<td>63%</td>
<td>75%</td>
<td>69%</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit match (NN,YY)</td>
<td>36%</td>
<td>43%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>Explicit mismatch (NY,YN)</td>
<td>8%</td>
<td>1%</td>
<td>1%</td>
<td>3%</td>
</tr>
<tr>
<td>Unclear mismatch (UY, UN)</td>
<td>37%</td>
<td>37%</td>
<td>31%</td>
<td>25%</td>
</tr>
<tr>
<td>NA mismatch (NaY, NaN)</td>
<td>19%</td>
<td>19%</td>
<td>63%</td>
<td>63%</td>
</tr>
</tbody>
</table>

most explicit mismatches for contact and financial information
mismatched expectations

collection of contact information

websites collect contact information without an account but participants don’t expect it

Yes — No
(Website – user) mismatch
mismatched expectations
sharing of contact information for other purposes

participants expect that contact information is shared with third parties for any reason but websites do not share it for non-core purposes
mismatched expectations

<table>
<thead>
<tr>
<th>Category</th>
<th>Collect without account</th>
<th>Collect with account</th>
<th>Share for core purpose</th>
<th>Share for other purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit match (NN,YY)</td>
<td>33%</td>
<td>89%</td>
<td>70%</td>
<td>23%</td>
</tr>
<tr>
<td>Explicit mismatch (NY,YN)</td>
<td>54%</td>
<td>5%</td>
<td>12%</td>
<td>65%</td>
</tr>
<tr>
<td>Unclear mismatch (UY, UN)</td>
<td>13%</td>
<td>6%</td>
<td>19%</td>
<td>13%</td>
</tr>
<tr>
<td>NA mismatch (NaY, NaN)</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Financial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit match (NN,YY)</td>
<td>20%</td>
<td>46%</td>
<td>41%</td>
<td>44%</td>
</tr>
<tr>
<td>Explicit mismatch (NY,YN)</td>
<td>43%</td>
<td>17%</td>
<td>28%</td>
<td>25%</td>
</tr>
<tr>
<td>Unclear mismatch (UY, UN)</td>
<td>12%</td>
<td>12%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>NA mismatch (NaY, NaN)</td>
<td>25%</td>
<td>25%</td>
<td>31%</td>
<td>31%</td>
</tr>
<tr>
<td>Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit match (NN,YY)</td>
<td>19%</td>
<td>18%</td>
<td>12%</td>
<td>20%</td>
</tr>
<tr>
<td>Explicit mismatch (NY,YN)</td>
<td>19%</td>
<td>20%</td>
<td>13%</td>
<td>11%</td>
</tr>
<tr>
<td>Unclear mismatch (UY, UN)</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>NA mismatch (NaY, NaN)</td>
<td>63%</td>
<td>63%</td>
<td>75%</td>
<td>69%</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit match (NN,YY)</td>
<td>36%</td>
<td>43%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>Explicit mismatch (NY,YN)</td>
<td>8%</td>
<td>1%</td>
<td>1%</td>
<td>3%</td>
</tr>
<tr>
<td>Unclear mismatch (UY, UN)</td>
<td>37%</td>
<td>37%</td>
<td>31%</td>
<td>25%</td>
</tr>
<tr>
<td>NA mismatch (NaY, NaN)</td>
<td>19%</td>
<td>19%</td>
<td>63%</td>
<td>63%</td>
</tr>
</tbody>
</table>

most explicit mismatches for contact and financial information

but policies are less clear for health and location
participants expect websites to permit deletion, but most websites don’t
mismatched expectations

summary

• few explicit mismatches but policies often unclear or silent on certain practices
• information collection without account often unexpected (contact, financial)
• participants assume that sharing is not limited to core purposes (e.g. also marketing)
• participants expect to be able to fully delete information, but most websites don’t allow it
limitations

• practices disclosed in privacy policy may not match service’s actual behavior
• online / MTurk study to elicit expectations
• additional practices may be of interest
• additional websites may be of interest
highlighting unexpected practices

<table>
<thead>
<tr>
<th>display in notice</th>
<th># practices</th>
<th>% reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>All practices</td>
<td>17</td>
<td>–</td>
</tr>
<tr>
<td>Mismatched practices only</td>
<td>11</td>
<td>35%</td>
</tr>
<tr>
<td>Unexpected practices only (Yes—No mismatch)</td>
<td>5</td>
<td>70%</td>
</tr>
</tbody>
</table>

potential reduction in information that users have to process in a layered or short notice
conclusions

- privacy expectations vs. preferences
- elicit expectations in surveys & compare with stated or actual data practices
- privacy expectations are affected by website type, privacy awareness, age, and experience with website
- opportunity for contextualizing and personalizing notices
- highlighting unexpected practices could reduce user burden and facilitate more informed privacy decision making

**Florian Schaub**  
fschaub@cmu.edu  
usableprivacy.org  

Carnegie Mellon University  
moving in the fall to

SCHOOL OF INFORMATION  
UNIVERSITY OF MICHIGAN