Are digital reference services color blind?

Pnina Shachaf

School of Library and Information Science, Indiana University, Bloomington
Shachaf@indiana.edu
Discrimination is:

- The act, practice, or an instance of discriminating categorically rather than individually (Merriam Webster)
- Behavior (not attitude or perception).
- Service discrimination: Reduced or enhanced online
Background

- Service discrimination against minorities
  - Rejection or poor services (Faegan, 1991).
  - Written requests for service (LaPiere, 1934; Glaser & Kahn, 2005).

- Digital reference services are provided in an increasingly multicultural environment.
  - Growing use of online services by minorities (Pew Internet & American Life Project, 2005).
  - Lack of empirical research on service discrimination online.
E-service: mixed research

- Virtual environment could increase subjective bias and give rise to inequality
  - People can become less self-aware and less likely to monitor their behavior online (e.g., Douglas & McGarty, 2001)
  - Trolling, vandalism, and other deviant online behaviors.
  - Discrimination & prejudice:
    - e.g., the lost email study found discrimination against Arabs (Bushman & Bonacci, 2004).

- Online discrimination
E-service: mixed research

- Virtual environment could also lessen subjective bias
  - Decreases social cues (Sproull & Keisler, 1986).
  - Eliminates blatant differences.
  - Members of target groups can hide their identity (Glaser & Kahn, 2005).

- Online equality
RQ

- Do e-services provide unbiased services to diverse user groups?
  - H1. All user groups will receive a response in a timely manner
  - H2. All user groups will receive the same level of reliable service
  - H3. All user groups will receive an equal level of courtesy
Method

- Unobtrusive
  - Reference (e.g., Hernon & McClure, 1986); Virtual reference (e.g., Arnold & Kaske, 2005; Stacy-Bates, 2003); Sensitive variables, such as ethnicity (Bushman & Bonacci, 2004)

- Information needs scenarios (10)
- Manipulated user names (10)
- Hotmail email accounts (10)
Data collection

- Pilot study
  - 23 ARL libraries, 138 transactions, 6 names, 6 Qs, 6 weeks

- Two follow-up studies
  - 88 public libraries, 352 transactions, 4 names, 4 Qs, 4 weeks
  - 54 academic libraries, 324 transactions, 6 names, 6Qs, 12 weeks

- Total
  - 155 libraries, 676 transactions, 10 names
Manipulation

Ahmed Ibrahim
Rosa Manuz
Latoya Johnson
Chang Su
Mary Anderson
Moshe Cohen

Pilot study & Academic libraries
Method (information need scenarios)

- **Week 1: Mary Anderson**
  Could you tell me the population of [city name] in 1963 and 1993?

- **Week 2: Moshe Cohen**
  Could you help me find information about [special collection topic]? Can you send me copies of articles on this topic?

- **Week 3: Ahmed Ibrahim**
  Can you send me by email a copy of the article “Free Indirect Discourse and Narrative Authority in Emma” by D.P. Gunn?

- **Week 4: Latoya Johnson**
  Can you tell me the title of [author]’s dissertation? He finished his degree at [institution name] in 1964. Do you have it in your library? How can I obtain a copy of this dissertation?

- **Week 5: Rosa Manuz**
  How did [sports team name] become the name for [institution name]’s sports teams? Can you refer me to a book or article that discusses it?

- **Week 6: Chang Su**
  Could you please send me a pdf copy of pp. 66-69 (ch.V) from Strunk and White's The Elements of Style?
Data analysis

- Content analysis using NVivo 2.0
- Coding scheme based on IFLA and RUSA guidelines (IFLA, 2005; RUSA, 2004a, 2004b)
- Inter-coder reliability
  - Academic - 92%, Cohen’s Kappa = .787
  - Public - 100%, Cohen’s Kappa = 1
- One-way ANOVA and cross-tabulations using SPSS
Pilot study

Responsiveness

counted number of days until reply

Weekends – Mary & Latoya
Responsiveness
counted number of days until reply

Level of service - Mary and Moshe vs. Latoya and Ahmed
Responsiveness

- ANOVA was significant, $F(5, 170) = 2.654, p = .025$.
- Significant pairwise differences:
  - between Ahmed and all other users, except for Mary and Latoya,
  - between Latoya and all other users, except for Mary and Ahmed.
- H1 was not supported; all users do not receive the same level of service in terms of responsiveness.
- Ahmed and Latoya are discriminated against.
Proportion of responses / answers

Pilot study

Ahmed and Latoya vs. Moshe and Mary
Pilot study

Answers (by type of Qs & user name)

Mary vs. Latoya and Ahmed
Pilot study

Closure

Rosa and Mary vs. Ahmed and Latoya
Reliability

- Users and all measures of reliability were found not to be significantly related:
  - Proportion of responses to the requests - Pearson $\chi^2 (5, N = 170) = 6.26, p = .282$, Cramer’s $V = .192$.
  - Use of librarian’s name - Pearson $\chi^2 (5, N = 138) = 1.683, p = .891$, Cramer’s $V = .110$.
  - Use of department’s name - Pearson $\chi^2 (5, N = 138) = 3.461, p = .629$, Cramer’s $V = .158$.

- H2 was supported; all users receive the same level of service in terms of reliability.
Ahmed
Users and all measures of courtesy were found not to be significantly related:

- Use of first name - Pearson $\chi^2 (5, N = 138) = 3.293, p = .655$, Cramer’s $V = .148$.
- Use of greetings - Pearson $\chi^2 (5, N = 138) = 1.35, p = .929$, Cramer’s $V = .099$.

H3 was supported; all users receive the same level of service in terms of courtesy.
Results

- Relationship between response time and users was not significant: F (5,318) = 1.969, p = .083
- H1 supported: all users received the same level of responsiveness

- Relationship between courtesy measures and users not significant
- H2 partially supported: Users received the same level of reliable service

- Relationship between reliability measures and users not significant
- H3 supported: Users received an equal level of courtesy

<table>
<thead>
<tr>
<th>Variable</th>
<th>Code</th>
<th>Pearson $\chi^2$</th>
<th>Cramer’s V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courtesy¹</td>
<td>First Name²</td>
<td>13.831*</td>
<td>.207*</td>
</tr>
<tr>
<td>Hello</td>
<td></td>
<td>3.956</td>
<td>.11</td>
</tr>
<tr>
<td>Reliability</td>
<td>Librarian’s name</td>
<td>8.348</td>
<td>.158</td>
</tr>
<tr>
<td></td>
<td>Department’s name</td>
<td>2.156</td>
<td>.082</td>
</tr>
<tr>
<td></td>
<td>Contact info.</td>
<td>6.254</td>
<td>.139</td>
</tr>
<tr>
<td></td>
<td>Completeness</td>
<td>1.533</td>
<td>.076</td>
</tr>
<tr>
<td></td>
<td>Accuracy³</td>
<td>3.356</td>
<td>.102</td>
</tr>
</tbody>
</table>

N=324, df=5

1. The frequencies for full name and honorifics per cell were less than five for some users; cross-tabulation in these cases could not be meaningful and therefore was not conducted.
2. Follow-up pair-wise comparisons using the Tukey’s HSD method was to control for type I error at the .05 level across all six comparisons. The significant pair-wise differences were between Chang Su and all other users.
Results: Hypothesis 1 (Responsiveness)

- Counted number of hours until reply
- **Not** significantly related to users \[F (3, 297) = 1.18, p = .318\], ethnicity \[F (1, 297) = .894, p = .345\], & gender \[F (1, 297) = .135, p = .714\].
- **H1 supported:** all users were treated with the **same level of responsiveness**
Results: Hypothesis 2 (Reliability)

- Coded answers as complete, partial, and / or accurate
- **Not** significantly related to users, gender, or ethnicity
- **H2 supported:** Users received the same level of reliable service
Results: Hypothesis 3 (Courtesy)

- Coded courtesy measures: i.e., hello, thanks
- **Not** significantly related to users, gender or ethnicity
- **H3 supported:**
  Users received an equal level of courtesy

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson $\chi^2$</td>
<td>Cramer's V</td>
</tr>
<tr>
<td>First Name</td>
<td>.431</td>
<td>.036</td>
</tr>
<tr>
<td>Full Name</td>
<td>.040</td>
<td>.011</td>
</tr>
<tr>
<td>Honorific</td>
<td>.325</td>
<td>.031</td>
</tr>
<tr>
<td>Last Name</td>
<td>1.9</td>
<td>.076</td>
</tr>
<tr>
<td>Thanks (op)</td>
<td>.354</td>
<td>.033</td>
</tr>
<tr>
<td>Hello</td>
<td>.296</td>
<td>.030</td>
</tr>
<tr>
<td>Thanks (cl)</td>
<td>.878</td>
<td>.054</td>
</tr>
<tr>
<td>Sincerely</td>
<td>.579</td>
<td>.042</td>
</tr>
</tbody>
</table>

N=352

* $p<.05$  ** $p<.01$
Results

All 3 hypotheses were supported in both follow-up studies.

Equitable level of service: responsiveness, reliability, and courtesy.

Only one of the 3 hypotheses was not supported in the pilot study.

Ahmed and Latoya are discriminated against: responsiveness.
Discussion

- Equitable e-services to diverse user groups
- Possible explanations
  1. Librarians core ethics
  2. ICT mediation
  3. Limitations
1. Librarians’ ethical values

- “...high level of service to all library users through... accurate, unbiased, and courteous responses to all requests.” (ALA code of ethics, 1995).

- Free and equal access is a core value of librarianship (e.g., Froehlich, 1997; Koehler & Pamberton, 2000; Gorman, 2000; Shachaf, 2005; Smith, 1997).
2. ICT mediation

- **Equalization** (e.g., Daily & Steiner, 1998). **Social presence is reduced; targets can hide their blatant differences, social and group identification cues; communication is more task oriented** (e.g., Sproull & Keisler, 1986).

- **ICT (email in particular) mediates cultural diversity challenges** (Shachaf, 2005; 2007; 2008).
3. Limitations

- Assumptions about librarians (e.g., demographics).
- Unobtrusiveness (consent from).
- Only 3 quality dimensions (complexity of interaction).
Conclusion

- Digital reference services are color blind.
  - ICT mediates diversity and reduce discrimination.
  - Librarians provide equal level of service to all user groups.
  - Need for future research.
Thank you!


Contact information:
Pnina Shachaf
shachaf@indiana.edu